

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 13:48:27 ; Search time 10429 Seconds  
(without alignments)  
11580.846 Million cell updates/sec

Title: US-10-016-725-15

Perfect score: 4150

Sequence: 1 aggaattccatttaaat.....tgagttccagctggccc 4150

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pin.\*

35: em.htg.rod.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htg.hum.\*

40: em.htg.mus.\*

41: em.htg.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4150	100.0	4150	9	AF320307	Homo sapi
c	4120.2	99.3	159351	2	AC018783	Homo sapi
3	4109.8	99.0	115278	2	AL139819	Human DNA
c	1992	48.0	185644	2	AL391723	Homo sapi
4	1932.8	46.6	1970	6	AX431388	Sequence
6	1448.2	34.9	179879	2	AC026883	Homo sapi
7	612.4	14.8	617	6	AX357255	Sequence
8	460	11.1	164201	9	AL157935	Human DNA
9	447.2	10.8	183386	2	AC119725	Homo sapi
10	433	10.4	192773	2	AL731547	Homo sapi
c	423.8	10.2	190508	9	AC005037	Homo sapi
12	415.2	10.0	44496	9	AC004760	Homo sapi
13	415.2	10.0	190821	9	AC006111	Homo sapi
c	408.8	9.9	183690	9	AC091180	Homo sapi
15	407.4	9.8	170528	9	AC083866	Homo sapi
16	404.2	9.8	217615	9	AC011005	Homo sapi
c	404.2	9.7	137289	9	AL161909	Human DNA
18	403	9.7	189179	2	AC090265	Homo sapi
c	402.8	9.7	181597	9	AC090515	Homo sapi
20	402.8	9.7	187734	9	AC025918	Homo sapi
21	394.2	9.5	179599	2	AC013278	Homo sapi
22	394.2	9.5	196894	2	AC007448	Homo sapi
23	392.2	9.5	155559	2	AL133336	Homo sapi
c	392.2	9.5	157581	2	AP005624	Homo sapi
25	391.6	9.4	171641	9	AC018512	Homo sapi
26	390	9.4	171555	2	AC087537	Homo sapi
c	390	9.4	181891	9	AC023356	Homo sapi
28	389.8	9.4	253178	2	AC025162	Homo sapi
29	388.4	9.4	172421	2	AC013279	Homo sapi
c	386.2	9.3	112364	9	AC109927	Homo sapi
31	386	9.3	151131	2	AC021945	Homo sapi
c	386	9.3	197092	2	AC036155	Homo sapi
33	385.8	9.3	147128	2	AC127512	Homo sapi
34	385.8	9.3	171936	2	AC120031	Homo sapi
35	385.8	9.3	191673	2	AL356743	Homo sapi
c	385.8	9.3	207418	9	AC090774	Homo sapi
37	385.8	9.3	208009	2	AC126365	Homo sapi
38	384.8	9.3	140529	2	AC011486	Homo sapi
c	384.2	9.3	167296	2	AC116346	Homo sapi
40	384	9.3	133475	9	AC006254	Homo sapi
41	382.4	9.2	188639	9	AC010997	Homo sapi
c	381.6	9.2	80775	9	AC068539	Homo sapi
43	381.2	9.2	150681	9	AC006011	Homo sapi
c	381.2	9.2	187064	9	AC011476	Homo sapi
45	381.2	9.2	204340	2	AC019238	Homo sapi

ALIGNMENTS

RESULT 1  
AF320307  
LOCUS Homo sapiens stearyl-CoA desaturase (SCD) gene, exon 1 and intron 1 and partial cds.  
DEFINITION AF320307  
ACCESSION AF320307.1 GI:14150490  
VERSION  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4150)  
AUTHORS Zhang L., Ge L., Tran, T., Stenn, K. and Prouty, S.M.  
TITLE Isolation and characterization of the human stearyl-CoA desaturase

gene promoter: requirement of a conserved CCAAT cis-element  
Biochem. J. 357 (Pt 1), 183-193 (2001)  
MEDLINE 21308466  
PUBMED 11415448  
REFERENCE 2 (bases 1 to 4150)  
AUTHORS Zhang, L., Ge, L., Tran, T., Stenn, K. and Prouty, S.M.  
Direct Submission  
TITLE Submitted (09-NOV-2000) Skin Biology TRC, Johnson and Johnson,  
JOURNAL CPWW, 199 Grandview Road, Skillman, NJ 08558, USA  
FEATURES  
source  
1. 4150  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/tissue\_type="scalp skin keratinocytes; liver; hair  
follicles; brain"  
3602..>4150  
/gene="SCD"  
CAAT\_signal  
3602..3606  
/gene="SCD"  
TATA\_signal  
3869..3872  
/gene="SCD"  
TATA\_signal  
3924..3927  
/gene="SCD"  
mRNA  
3958..>4129  
/gene="SCD"  
/product="stearoyl-CoA desaturase"  
/note="3602"  
3958..4129  
/gene="SCD"  
/number=1  
3958..4102  
/gene="SCD"  
4103..>4129  
/gene="SCD"  
/codon\_start=1  
/product="stearoyl-CoA desaturase"  
/protein\_id="AAK54510.1"  
/db\_xref="GI:14150491"  
/translation="MPAHLQDD"  
4130..>4150  
/gene="SCD"  
/number=1  
BASE COUNT 988 a 1076 c 1065 g 1021 t  
ORIGIN  
Query Match 100.0%; Score 4150; DB 9; Length 4150;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGAATCATCAATTAATCAATCAATTAATGGCTTTAGTATATTCACAGGTTGTGC 60  
Db 1 AGGAATCATCAATTAATCAATCAATTAATGGCTTTAGTATATTCACAGGTTGTGC 60  
QY 61 ATCCATCACATCCATTTAGAACAGTTTATTACTCCAAAATAAACCCTGCATTCCTT 120  
Db 61 ATCCATCACATCCATTTAGAACAGTTTATTACTCCAAAATAAACCCTGCATTCCTT 120  
QY 121 AGCCATCACCCCAACATCCCATCCCTCCCAAGCCCTGGGCAACCAACATCTAC 180  
Db 121 AGCCATCACCCCAACATCCCATCCCTCCCAAGCCCTGGGCAACCAACATCTAC 180  
QY 181 TTCTGCTCTATAAATTTGCCAATTCGGACATTCATATAAATGGNAGCAACACAT 240  
Db 181 TTCTGCTCTATAAATTTGCCAATTCGGACATTCATATAAATGGNAGCAACACAT 240  
QY 241 GTGAGACTTTGTGACTGGCTCTTCACTTAGCATTCCTATTTTAAAGGCTCATTTGTTA 300  
Db 241 GTGAGACTTTGTGACTGGCTCTTCACTTAGCATTCCTATTTTAAAGGCTCATTTGTTA 300  
QY 301 CAGTACTAGCAGTACTTCAATCTTTTATTCCTCAAAATGGTATTCACCTGTGTGGTAT 360  
Db 301 CAGTACTAGCAGTACTTCAATCTTTTATTCCTCAAAATGGTATTCACCTGTGTGGTAT 360

QY 361 CCATATCATATTATTAGAGACAGGTTCTCTCTGTCCACCAGGCTGGAGTGCAGTGGC 420  
Db 361 CCATATCATATTATTAGAGACAGGTTCTCTCTGTCTGTCCACCAGGCTGGAGTGCAGTGGC 420  
QY 421 ACAATCATAGCTCACTGTAACTCAAACTCTGGGCTCAAGTGAATCCCTACCTACGAC 480  
Db 421 ACAATCATAGCTCACTGTAACTCAAACTCTGGGCTCAAGTGAATCCCTACCTACGAC 480  
QY 481 TCCAGAGTAGCTAGGACTACAGGCACACACAGCCATACCTGCTAATTTTTTTTTTAAAT 540  
Db 481 TCCAGAGTAGCTAGGACTACAGGCACACACAGCCATACCTGCTAATTTTTTTTTTAAAT 540  
QY 541 TTTCATTTTATGTATTCATTTTCTTTTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGT 600  
Db 541 TTTCATTTTATGTATTCATTTTCTTTTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGT 600  
QY 601 ACTTTGTACCCAGGCTGGAGGCGAGTGGCATGGTGGAGCTGACAGCTGACGCCCTTGCT 660  
Db 601 ACTTTGTACCCAGGCTGGAGGCGAGTGGCATGGTGGAGCTGACAGCTGACGCCCTTGCT 660  
QY 661 GGGCTCAAGTGAATCTCTCTGCTCAGCCCTCCCAAGTAGCTGGGACTACAAACACGCTGCA 720  
Db 661 GGGCTCAAGTGAATCTCTCTGCTCAGCCCTCCCAAGTAGCTGGGACTACAAACACGCTGCA 720  
QY 721 CCATGCTGGCTGATATTTTCTTTTCTTTTGAACAGGCTATCATCTCTGTGTGTGTGTGTGT 780  
Db 721 CCATGCTGGCTGATATTTTCTTTTCTTTTGAACAGGCTATCATCTCTGTGTGTGTGTGTGT 780  
QY 781 GTACAGTGGCGTAATAATAGCTCACTGAGCTCCCTCTCTGGGCTCAAGCAATCCGCTG 840  
Db 781 GTACAGTGGCGTAATAATAGCTCACTGAGCTCCCTCTCTGGGCTCAAGCAATCCGCTG 840  
QY 841 GCCTCAGCATCCTGAGTAGCTGGGACTACAGGCTGTGCCACAGGCCAGCTAAGTTTT 900  
Db 841 GCCTCAGCATCCTGAGTAGCTGGGACTACAGGCTGTGCCACAGGCCAGCTAAGTTTT 900  
QY 901 AAAAAATGATTTTGGTATAGAGGAGTCTTGTGTATGTGTGTGTGTGTGTGTGTGTGTGT 960  
Db 901 AAAAAATGATTTTGGTATAGAGGAGTCTTGTGTATGTGTGTGTGTGTGTGTGTGTGTGT 960  
QY 961 TTGAGACAAGGCTCACTATGTGTCATGATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1020  
Db 961 TTGAGACAAGGCTCACTATGTGTCATGATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1020  
QY 1021 TCCTATCTGTTTCATTAGTACAGTGTGACAGCATTTAGGTTGTGTGTGTGTGTGTGTGTGTGT 1080  
Db 1021 TCCTATCTGTTTCATTAGTACAGTGTGACAGCATTTAGGTTGTGTGTGTGTGTGTGTGTGTGT 1080  
QY 1081 TGAATATACCTCCAGTGAATATTCATGTATACATTTGTGTGGCATATGTTTTCATTTCT 1140  
Db 1081 TGAATATACCTCCAGTGAATATTCATGTATACATTTGTGTGGCATATGTTTTCATTTCT 1140  
QY 1141 GTTGGGTTTATCTAGGAGTGGAAATGCTGGATCCCGGGTAATTTTGTACAGGCGAG 1200  
Db 1141 GTTGGGTTTATCTAGGAGTGGAAATGCTGGATCCCGGGTAATTTTGTACAGGCGAG 1200  
QY 1201 TTCAGGGGAAGAAACCTTGGGAAATGAAGCATGTTTAGAAATCAGCAAGTGCAGGG 1260  
Db 1201 TTCAGGGGAAGAAACCTTGGGAAATGAAGCATGTTTAGAAATCAGCAAGTGCAGGG 1260  
QY 1261 GTTTTTCGGAGTTTATTTATATTTCTGTGACAAATGTGCGAGTTTGTATGAAGATACAAG 1320  
Db 1261 GTTTTTCGGAGTTTATTTATATTTCTGTGACAAATGTGCGAGTTTGTATGAAGATACAAG 1320  
QY 1321 TTATCTAAGTGAAGTGAATTAAGGCTGGAAATAGGGGCTTCAGAGTAAAAATCATGA 1380  
Db 1321 TTATCTAAGTGAAGTGAATTAAGGCTGGAAATAGGGGCTTCAGAGTAAAAATCATGA 1380  
QY 1381 AGCACTTGAATCAACCAAAATTAAGGAGCTTGGCTGTGAACAAATAATAAAAAATCAAA 1440  
Db 1381 AGCACTTGAATCAACCAAAATTAAGGAGCTTGGCTGTGAACAAATAATAAAAAATCAAA 1440

QY	1441	TTTTTTTTTTTTTTTTTTGAGAAAGTCTTTGCHCTTTACCTGGCTGGAGCGCAGTGGTG	1501
Db	1441	TTTTTTTTTTTTTTTTTTGAGAAAGTCTTTGCHCTTTACCTGGCTGGAGCGCAGTGGTG	1501
QY	1501	TGATCTCAGCTCACTCAACTTTTCGCTCCCGGGTTCAAGCAATTCCTCCGCTCAGCCT	1560
Db	1501	TGATCTCAGCTCACTCAACTTTTCGCTCCCGGGTTCAAGCAATTCCTCCGCTCAGCCT	1560
QY	1561	CCCAAGTAGCTGGGACTACAGGCACCTCCCACTATCCCACTGCGCCGCTTATTTTATTTT	1620
Db	1561	CCCAAGTAGCTGGGACTACAGGCACCTCCCACTATCCCACTGCGCCGCTTATTTTATTTT	1620
QY	1621	TAGAGATGGGATTTTCACTTTGTTGGCCAAAGCTGGTCTCAAACTTTTTCGCTCATAAITG	1680
Db	1621	TAGAGATGGGATTTTCACTTTGTTGGCCAAAGCTGGTCTCAAACTTTTTCGCTCATAAITG	1680
QY	1681	TTGTAACATTTGTTCTCTTTGCTGAGGTAGGGCCCCAGACCAAAAAAATAAATCTTAG	1740
Db	1681	TTGTAACATTTGTTCTCTTTGCTGAGGTAGGGCCCCAGACCAAAAAAATAAATCTTAG	1740
QY	1741	AAATCCAAATCAGTGTGTGGTTGTGACCACTGTCACTTGAGAACACACAGTGTGACAGGGC	1800
Db	1741	AAATCCAAATCAGTGTGTGGTTGTGACCACTGTCACTTGAGAACACACAGTGTGACAGGGC	1800
QY	1801	CTCAGGAGTAGAGGTGATCTCTGCTCGAAAGAGAAAATAGAAATGAAATATTCCTCCGGGCC	1860
Db	1801	CTCAGGAGTAGAGGTGATCTCTGCTCGAAAGAGAAAATAGAAATGAAATATTCCTCCGGGCC	1860
QY	1861	AGCGTGTGGTCAATGCTGTAAATCCAGCACCTTTGGGAGGCCAAGCATGTGGATCAC	1920
Db	1861	AGCGTGTGGTCAATGCTGTAAATCCAGCACCTTTGGGAGGCCAAGCATGTGGATCAC	1920
QY	1921	CTCAGGTCTAGGAGTTCAAAACCAGCTGGCCAAACATGCTGAAACCCCGTCTCTACTAAAA	1980
Db	1921	CTCAGGTCTAGGAGTTCAAAACCAGCTGGCCAAACATGCTGAAACCCCGTCTCTACTAAAA	1980
QY	1981	ATACAAAAATTAGCTAAGTGTGGTGGCCATGCTGTAAATCCAGCTACTTTGGGAGGGT	2040
Db	1981	ATACAAAAATTAGCTAAGTGTGGTGGCCATGCTGTAAATCCAGCTACTTTGGGAGGGT	2040
QY	2041	GAGGCAGGAGAAATTTCTGAACCCGGGAGGAGAGGTTCAGTCAAGCGAGATCACACCA	2100
Db	2041	GAGGCAGGAGAAATTTCTGAACCCGGGAGGAGAGGTTCAGTCAAGCGAGATCACACCA	2100
QY	2101	CTGCATCCAGCCTGGGGAGAGACGAGACTTCCTCTCAAAAAACAACAAAAACAAGA	2160
Db	2101	CTGCATCCAGCCTGGGGAGAGACGAGACTTCCTCTCAAAAAACAACAAAAACAAGA	2160
QY	2161	ATTAAACAAATTAGACATTTGCAGACAGACCTGAGGGGTCAGACCACCTACAGATTTC	2220
Db	2161	ATTAAACAAATTAGACATTTGCAGACAGACCTGAGGGGTCAGACCACCTACAGATTTC	2220
QY	2221	TGTGCCACATGCCAAGTACTTCTGAGGCATGACTGGATGAGCTGTCACACATCTGAAATCA	2280
Db	2221	TGTGCCACATGCCAAGTACTTCTGAGGCATGACTGGATGAGCTGTCACACATCTGAAATCA	2280
QY	2281	TCCAGTCTGTTTCAGAACTTTACACCCGACAGGAGCCAGGACTGGAATGCAAGTCTCCT	2340
Db	2281	TCCAGTCTGTTTCAGAACTTTACACCCGACAGGAGCCAGGACTGGAATGCAAGTCTCCT	2340
QY	2341	GGTCACTGCCACAGAGTTGGCCTTGACCTTGAGACACAGTGGCCAAACAAAGGAGCTGCTT	2400
Db	2341	GGTCACTGCCACAGAGTTGGCCTTGACCTTGAGACACAGTGGCCAAACAAAGGAGCTGCTT	2400
QY	2401	AGTCTACCTCCGAGGAATCCCAAGTGGTGTGTCTTCCTGGGAAGTGAATCATTTGGCGCAG	2460
Db	2401	AGTCTACCTCCGAGGAATCCCAAGTGGTGTGTCTTCCTGGGAAGTGAATCATTTGGCGCAG	2460
QY	2461	CACCTCCGATTTTCTCCTTTCCCAAGGGAAGGATCTTAGGGCAGTATTTGGGAAAGACA	2520
Db	2461	CACCTCCGATTTTCTCCTTTCCCAAGGGAAGGATCTTAGGGCAGTATTTGGGAAAGACA	2520
QY	2521	TGGGCATGGAAGGACACCGGTTGAATGCAATAGCCTGGCTGGTCTGAGCTTCATGGTAA	2580

D	b	2521	TGGGCATGGGAAGACACCGGGTGAATGACATACGCTGCTGGTCTTGAGAGCTCTCATGTAA	2581
Q	y	2581	GGCTTCCTACAGACACAGAAAGATGGGCGACAGAGGACAGATACGTAGGGTCAGAGCATC	2640
D	b	2581	GGCTCCTACAGACAGGAAAGATGGGCGCACAGGCACAGATCAGTAGGGTCAGAGCATC	2640
Q	y	2641	TCAGGGACGAGGGGCAATATGTCCTTGACAGAGGATTAAGAGCTTGGGCTCTCATATGGT	2700
D	b	2641	TCAGGGACGAGGGCAATATGTCCTTGACAGAGGATTAAGAGCTTGGGCTCTCATATGGT	2700
Q	y	2701	GTTTCTGGGCTCAACTGCCAGCTCCGTCACCTTACCTGCTTGCTGTGACCATGGCAAGTTA	2760
D	b	2701	GTTTCTGGGCTCAACTGCCAGCTCCGTCACCTTACCTGCTTGCTGTGACCATGGGCAAGTTA	2760
Q	y	2761	TTCCATCTCTCCATATCTCTTCTCTACPTTTTAAATGGAATAATGGGGTACCCACCTCC	2820
D	b	2761	TTCCATCTCTCCATATCTCTTCTCTACPTTTTAAATGGAATAATGGGGTACCCACCTCC	2820
Q	y	2821	CAGGGTCACAGAGAGGCTTACAGAAACGATCTTGTCATTTGGCTGTCAGTATATATTC	2880
D	b	2821	CAGGGTCACAGAGAGGCTTACAGAAACGATCTTGTCATTTGGCTGTCAGTATATATTC	2880
Q	y	2881	AATACTGCCAGCTATCTTATTCCACATCCAAAGCCCTTCGCCCTGCTGGTGAGAAA	2940
D	b	2881	AATACTGCCAGCTATCTTATTCCACATCCAAAGCCCTTCGCCCTGCTGGTGAGAAA	2940
Q	y	2941	CACATGTCAGTGTTTCCCTGACGGTTTCACAAAGAGATTCCTCAATTTGGCTGTCAGTATATTC	3000
D	b	2941	CACATGTCAGTGTTTCCCTGACGGTTTCACAAAGAGATTCCTCAATTTGGCTGTCAGTATATTC	3000
Q	y	3001	TCTGAAGAATCTCCAAACATCCGACAGCATCTCTGAGGCGCGGGCTTGGGATGGGAC	3060
D	b	3001	TCTGAAGAATCTCCAAACATCCGACAGCATCTCTGAGGCGCGGGCTTGGGATGGGAC	3060
Q	y	3061	TGCGCGCCGGGTCCTGTAACAGATGCTGTGCGCGAGGCAACACACAGCCAGCGCTG	3120
D	b	3061	TGCGCGCCGGGTCCTGTAACAGATGCTGTGCGCGAGGCAACACACAGCCAGCGCTG	3120
Q	y	3121	TGTGTGCGCGCGAGTCCGGTCCGGGTGAGCAGCGCTGCTGTTGGCGGGGCG	3180
D	b	3121	TGTGTGCGCGCGAGTCCGGTCCGGGTGAGCAGCGCTGCTGTTGGCGGGGCG	3180
Q	y	3181	AGAGCCATTGTTGCGAGGCGTACCGAGGCCCCCGCTCGCCCCGGGAGGAGCGGGCT	3240
D	b	3181	AGAGCCATTGTTGCGAGGCGTACCGAGGCCCCCGCTCGCCCCGGGAGGAGCGGGCT	3240
Q	y	3241	TCCGCGCTCCCAAGCTCCAGATCTCGGGTGGCTGCCACGTCCTCGCCAGCGCGCTG	3300
D	b	3241	TCCGCGCTCCCAAGCTCCAGATCTCGGGTGGCTGCCACGTCCTCGCCAGCGCGCTG	3300
Q	y	3301	GGGGGACGGGAACAGCGGACGGAGATGTAGTGTGGGCGCCCCCGAGGGTTTCAACACT	3360
D	b	3301	GGGGGACGGGAACAGCGGACGGAGATGTAGTGTGGGCGCCCCCGAGGGTTTCAACACT	3360
Q	y	3361	GTTTCTTGAGAACTTCCCGAGTGCACCCACCGTGTCTCCGTGTGCCCGAGGGCGGCT	3420
D	b	3361	GTTTCTTGAGAACTTCCCGAGTGCACCCACCGTGTCTCCGTGTGCCCGAGGGCGGCT	3420
Q	y	3421	CCTGGGCTAGGCTCCGCGCCCCCAGCCCAACCGGGTCCCCCAGCCCCCTTCCAGAGAGAA	3480
D	b	3421	CCTGGGCTAGGCTCCGCGCCCCCAGCCCAACCGGGTCCCCCAGCCCCCTTCCAGAGAGAA	3480
Q	y	3481	GCTCCCGACGGGGATGCCGGGACGAGGCCACCGCGGGGTGGGAAGAGAAGCTGAGAAGG	3540
D	b	3481	GCTCCCGACGGGGATGCCGGGACGAGGCCACCGCGGGGTGGGAAGAGAAGCTGAGAAGG	3540
Q	y	3541	AGAAACAGAGGGAGGGGACGAGGAGCTGGCGCACAGGGACACGATTTGCCCGCA	3600
D	b	3541	AGAAACAGAGGGAGGGGACGAGGAGCTGGCGCACAGGGACACGATTTGCCCGCA	3600
Q	y	3601	GCCAAATGGCAACGGCAGGACGAGGTGGCACCAAAATTCCTTTCGGGCCAATGACGAGCCGA	3660





```

misc_feature /note="assembly_fragment"
21924..25747
/note="assembly_fragment"
clone_end:SP6
vector_side:right
25848..31971
/note="assembly_fragment"
clone_end:T7
vector_side:left
32072..44449
/note="assembly_fragment"
44550..64978
/note="assembly_fragment"
65079..91491
/note="assembly_fragment"
91592..121937
/note="assembly_fragment"
122038..159351
/note="assembly_fragment"
BASE COUNT 44283 a 34858 c 35376 g 43731 t 1103 others
ORIGIN

Query Match 99.3%; Score 4120.2; DB-2; Length 159351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4144; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 3 GAATTCATCCATTAAATCATACAAATTTAATGGCTTTTAGTATATTCACAGGTTGTGCAT 62
Db 59390 GAATTCATCCATTAAATCATACAAATTTAATGGCTTTTAGTATATTCACAGGTTGTGCAT 59331
Qy 63 CCATCACAATCCATTTAGACAGTTTATTTACTCCAAAAATAAACCCCTGCATTCCTTAG 122
Db 59330 CCATCACAATCCATTTAGACAGTTTATTTACTCCAAAAATAAACCCCTGCATTCCTTAG 59271
Qy 123 CCATCACCCTCCACATCCTCCATCCTCTCCAAAGCCCTGGGCAACCAACCAATCTACT 182
Db 59270 CCATCACCCTCCACATCCTCCATCCTCTCCAAAGCCCTGGGCAACCAACCAATCTACT 59211
Qy 183 TCTGCTCTATAAATTCGCCAATTCGACATTTTCATATAATGAAGCAACACACATGT 242
Db 59210 TCTGCTCTATAAATTCGCCAATTCGACATTTTCATATAATGAAGCAACACACATGT 59151
Qy 243 GAGACTTTGTGACGTGGCTTTCACITTAGCATTTCTATTTTAAAGGCTCAATTTACTACA 302
Db 59150 GAGACTTTGTGACGTGGCTTTCACITTAGCATTTCTATTTTAAAGGCTCAATTTACTACA 59091
Qy 303 GTACTTAGCAGTACTTCAATCTTTTATTCCTCAATGGTATTCACGTGTGGGTATCC 362
Db 59090 GTACTTAGCAGTACTTCAATCTTTTATTCCTCAATGGTATTCACGTGTGGGTATCC 59031
Qy 363 CATATCATATTTATAGACAGGTTCTCACCTGTCCACCCAGGCTGGAGTGCAGTGGCAC 422
Db 59030 CATATCATATTTATAGACAGGTTCTCACCTGTCCACCCAGGCTGGAGTGCAGTGGCAC 58971
Qy 423 AATCATAGCTCAGTAACTCAACTCCTGGGCTCAAGTGATCCTACTACCTCAGCCTC 482
Db 58970 AATCATAGCTCAGTAACTCAACTCCTGGGCTCAAGTGATCCTACTACCTCAGCCTC 58911
Qy 483 CAGAGTAGCTAGGACTACAGGCACACAGCCATACCTGGCTAATTTTTTTTAAATTT 542
Db 58910 CAGAGTAGCTAGGACTACAGGCACACAGCCATACCTGGCTAATTTTTTTTAAATTT 58851
Qy 543 TCATTTATGATTCATTTCTTTCTTTTGTGTTGTTGTTTGTGATAGGCTCTCAC 602
Db 58850 TCATTTATGATTCATTTCTTTCTTTTGTGTTGTTGTTTGTGATAGGCTCTCAC 58791
Qy 603 TTTGTTACCCAGGCTGGAGGCACTGGCATGGTGACAGCTGAGCAGCCTTGACTTTCCTGG 662
Db 58790 TTTGTTACCCAGGCTGGAGGCACTGGCATGGTGACAGCTGAGCAGCCTTGACTTTCCTGG 58731
Qy 663 GCTCAAGTAGTCCCTCCCTCAGCCTCCCAAGTAGCTGGGACTACAAACACAGTGTCAAC 722
Db 58730 GCTCAAGTAGTCCCTCCCTCAGCCTCCCAAGTAGCTGGGACTACAAACACAGTGTCAAC 58671

```

```

Qy 723 ATGCGCTGGCTGATATATTTTTTTTGAACAGGGTATCACTCTGTGTCGCCAGGCTGGAGT 782
Db 58670 ATGCGCTGGCTGATATATTTTTTTTGAACAGGGTATCACTCTGTGTCGCCAGGCTGGAGT 58611
Qy 783 ACAGTGGCGTAATAATAGTCACTGAGCCTCCCTCCCTGGGCTCAAGCAATCCGCTGGC 842
Db 58610 ACAGTGGCGTAATAATAGTCACTGAGCCTCCCTCCCTGGGCTCAAGCAATCCGCTGGC 58551
Qy 843 CTCAGCATCCTGAGTAGCTGGGACTACAGGCTGTGCCACACAGGCCAGCTAAGTTTTAA 902
Db 58550 CTCAGCATCCTGAGTAGCTGGGACTACAGGCTGTGCCACACAGGCCAGCTAAGTTTTAA 58491
Qy 903 AAAATGATTTTTGGTATAGAGAGGCTCTTGTCTATGTTGTCTAGGCTGTATTTTATTTGTT 962
Db 58490 AAAATGATTTTTGGTATAGAGAGGCTCTTGTCTATGTTGTCTAGGCTGTATTTTATTTGTT 58431
Qy 963 GAGACAAGGCTCACTATGTTGCCATGATCCCCACCTCCACTTCCCAAGAGTGTCAATC 1022
Db 58430 GAGACAAGGCTCACTATGTTGCCATGATCCCCACCTCCACTTCCCAAGAGTGTCAATC 58371
Qy 1023 TTATCTGTTTCATTAGTTCAGTTGACAGACATTTAGTGTGTTTCCACTTTTGGACCATATG 1082
Db 58370 TTATCTGTTTCATTAGTTCAGTTGACAGACATTTAGTGTGTTTCCACTTTTGGACCATATG 58311
Qy 1083 AATAATACCTCAGTGAATTTATGATATACATTTGTTGTGGGCAATGTTTTCATTTCTGT 1142
Db 58310 AATAATACCTCAGTGAATTTATGATATACATTTGTTGTGGGCAATGTTTTCATTTCTGT 58251
Qy 1143 TGGGTTTATATCTAGAGTGGAAATTCGATCCCGGTAATATTTTGACAGGACAGATT 1202
Db 58250 TGGGTTTATATCTAGAGTGGAAATTCGATCCCGGTAATATTTTGACAGGACAGATT 58191
Qy 1203 CAGGGGAAGAAAACTTGGGAAATGAAGCATGTTTAGAAATCAGCAAGAGTGCAGGGGT 1262
Db 58190 CAGGGGAAGAAAACTTGGGAAATGAAGCATGTTTAGAAATCAGCAAGAGTGCAGGGGT 58131
Qy 1263 TTTTCGGAGTTTATTTTATATATTTCTGTGACAAATGTCAGTTTGTATGAAGATACAAGTT 1322
Db 58130 TTTTCGGAGTTTATTTTATATATTTCTGTGACAAATGTCAGTTTGTATGAAGATACAAGTT 58071
Qy 1323 ATACTAAGTGAGAACTGAGAAATTAAGGCTGGAATTAAGGGGTTTCAGAGTAAAAATCATGAAG 1382
Db 58070 ATACTAAGTGAGAACTGAGAAATTAAGGCTGGAATTAAGGGGTTTCAGAGTAAAAATCATGAAG 58011
Qy 1383 CACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAAAATAATAAAAAATCAAAAT 1442
Db 58010 CACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAAAATAATAAAAAATCAAAAT 57951
Qy 1443 TTTTTCCTTTTTCAGAAAGAGTCTTGTCTTTTCCACCTGGCTGGAGGCGAGTGGTGTG 1502
Db 57950 TTTTTCCTTTTTCAGAAAGAGTCTTGTCTTTTCCACCTGGCTGGAGGCGAGTGGTGTG 57891
Qy 1503 ATCTCAGCTCACTGCAACTTTTCGCCCTCCCGGTTTCAAGCAATTCCTCTGCTTCAGCCTCC 1562
Db 57890 ATCTCAGCTCACTGCAACTTTTCGCCCTCCCGGTTTCAAGCAATTCCTCTGCTTCAGCCTCC 57831
Qy 1563 CAAGTAGCTGGGACTACAGGCACTTCCCAACATGCCAGTCCAGCTGATTTTGTATTTTGTAGTA 1622
Db 57830 CAAGTAGCTGGGACTACAGGCACTTCCCAACATGCCAGTCCAGCTGATTTTGTATTTTGTAGTA 57771
Qy 1623 GAGATGGGATTTCACTTTGTTGGCAAGCTGGTCTCAAACTTTTGTCTGTATATTTGTT 1682
Db 57770 GAGATGGGATTTCACTTTGTTGGCAAGCTGGTCTCAAACTTTTGTCTGTATATTTGTT 57711
Qy 1633 GTAACATGTTTCCTTTTGTCTGAGTAGGGCCCCCAGACACAAAAATAATACTTAGAA 1742
Db 57710 GTAACATGTTTCCTTTTGTCTGAGTAGGGCCCCCAGACACAAAAATAATACTTAGAA 57651
Qy 1743 TCCAAATCAGTGTGTGTTGTTGACCACTGTCACTTTGAGAACACACAGTGTGACCAAGGCGCT 1802
Db 57650 TCCAAATCAGTGTGTGTTGTTGACCACTGTCACTTTGAGAACACACAGTGTGACCAAGGCGCT 57591

```

QY	1803	CAGGAGT	AGAGTGATCTCTGCTGCAAAAGAGAAATAGAAATAGAAATATCTCTCGGCGCCAG	1862
Db	57590	CAGGAGT	AGAGTGATCTCTGCTGCAAAAGAGAAATAGAAATAGAAATATCTCTCGGCGCCAG	57531
QY	1863	CGCTGGT	GGCTCATGCCCTGTAATCCAGCACTTTTGGGAGGCCAAGGCATGTGGATCACCT	1922
Db	57530	CGCTGGT	GGCTCATGCCCTGTAATCCAGCACTTTTGGGAGGCCAAGGCATGTGGATCACCT	57471
QY	1923	GAGGTC	AGAGGTTCAAAACCAAGCCCTGGCCAAACATGGTGAACCCCGCTCTCTACTAAAAAT	1982
Db	57470	GAGGTC	AGAGGTTCAAAACCAAGCCCTGGCCAAACATGGTGAACCCCGCTCTCTACTAAAAAT	57411
QY	1983	ACAAAAAAT	TAGCTAAGTGTGGTGGCGCATGCCTGTATATCCAGACTACTTGGGAGGGTGA	2042
Db	57410	ACAAAAAAT	TAGCTAAGTGTGGTGGCGCATGCCTGTATATCCAGACTACTTGGGAGGGTGA	57351
QY	2043	GGCAGG	AAATTTCTTTGAACCCGGGAGGACAGAGTTGCAAGTGAAGCGAGATCACACCACT	2102
Db	57350	GGCAGG	AAATTTCTTTGAACCCGGGAGGACAGAGTTGCAAGTGAAGCGAGATCACACCACT	57291
QY	2103	GCATCC	AGCCCTGGGGAGAGCGAGACTTTCCTCTCAAAAAACAACAAAAAGAAAT	2162
Db	57290	GCATCC	AGCCCTGGGGAGAGCGAGACTTTCCTCTCAAAAAACAACAAAAAGAAAT	57231
QY	2163	TAAGCA	ATTAGACATTCACAGAGAGAACTGAAAGGGGTACAGCACGTCACAGATTTCTG	222
Db	57230	TAAGCA	ATTAGACATTCACAGAGAGAACTGAAAGGGGTACAGCACGTCACAGATTTCTG	57171
QY	2223	TGCCAC	ATGCCAAGTACTTCTGAGCATGACTGGATGAGCTGTCCACATCTGAAATCATC	2282
Db	57170	TGCCAC	ATGCCAAGTACTTCTGAGCATGACTGGATGAGCTGTCCACATCTGAAATCATC	57111
QY	2283	CAGTCT	TGTTCCAGAACTTTCACCCGACAGGGAGCCAGGACTGGAATGCAGTCTCCTGG	2342
Db	57110	CAGTCT	TGTTCCAGAACTTTCACCCGACAGGGAGCCAGGACTGGAATGCAGTCTCCTGG	57051
QY	2343	TCATG	GCCAGAGTGTGGCCCTTGACCCCTGAGACCAAGTGGCCAAACAAAGGAGCTGCTTAG	2402
Db	57050	TCATG	GCCAGAGTGTGGCCCTTGACCCCTGAGACCAAGTGGCCAAACAAAGGAGCTGCTTAG	56992
QY	2403	TCTAC	CTCCCAAGGAAATCCCAAGTGCTTGCTTCTCTGGGAAGTCAATCATTTGGCGGACCA	2462
Db	56991	TCTAC	CTCCCAAGGAAATCCCAAGTGCTTGCTTCTCTGGGAAGTCAATCATTTGGCGGACCA	56932
QY	2463	CTCGT	ATTTTCCTCTTCCACGGGGAAGATCCTAGGGCAGTATTTGGGAAAGACATG	2522
Db	56931	CTCGT	ATTTTCCTCTTCCACGGGGAAGATCCTAGGGCAGTATTTGGGAAAGACATG	56872
QY	2523	GGCAT	GGAAGACACCGGGTGAATGATCATAGCTGCCCTGGTCTCAGCTCTCATGGTAAGG	2582
Db	56871	GGCAT	GGAAGACACCGGGTGAATGATCATAGCTGCCCTGGTCTCAGCTCTCATGGTAAGG	56812
QY	2583	CTCCT	TACAGACACGGAAAAGATGGGGCCACAGGGACAGATCATGAGGTCAGAGCATCTC	2642
Db	56811	CTCCT	TACAGACACGGAAAAGATGGGGCCACAGGGACAGATCATGAGGTCAGAGCATCTC	56752
QY	2643	AGGAC	CCAGGGCAATATGTCCTGACAGGGATTAAGACCTTGGGCTCTCATATGGTGT	2702
Db	56751	AGGAC	CCAGGGCAATATGTCCTGACAGGGATTAAGACCTTGGGCTCTCATATGGTGT	56692
QY	2703	TTCTG	GGGCTCAACTGCCAGCTCCGCTCACTTACTGGTGTCTGACCATTGGGCAAGTTATT	2762
Db	56691	TTCTG	GGGCTCAACTGCCAGCTCCGCTCACTTACTGGTGTCTGACCATTGGGCAAGTTATT	56632
QY	2763	CCATCT	CTCCATATCTTTTCCCTCACTTTTAAATGGAATTAATGGGGTACCCACCTCCCA	2822
Db	56631	CCATCT	CTCCATATCTTTTCCCTCACTTTTAAATGGAATTAATGGGGTACCCACCTCCCA	56572
QY	2823	GGGTC	ACAGAGGCTTACAGAAAAACGATTCTTGTGAATTTGGCTTGCAGTAATAATTCAA	2882
Db	56571	GGGTC	ACAGAGGCTTACAGAAAAACGATTCTTGTGAATTTGGCTTGCAGTAATAATTCAA	56512
QY	2883	TACCT	GCCAGCTATTCCTTATTCACATPCCAAAGCCCTTTCCGCTCTGCTCGGGTAAAAACA	2942

Db	56511	TACCTGCCAGCTATTTCTTATTCATATCAATCAAGCCCTTTCGCTGTGCTGGGTGAAAACA	56457
QY	2943	CATGTCAAGTGTTCCTTGACGGTTTCCACAAGAAGATTCCAAAAATTACAACTGCCAGTC	3002
Db	56451	CATGTCAAGTGTTCCTTGACGGTTTCCACAAGAAGATTCCAAAAATTACAACTGCCAGTC	56392
QY	3003	TGAAGAAATCCAAAACATCCCGCAGCAATCTGAGGCGCGGGCTTGGGATGGACATG	3062
Db	56391	TGAAGAAATCCAAAACATCCCGCAGCAATCTGAGGCGCGGGCTTGGGATGGACATG	56332
QY	3063	CCGCGCCGGGTCTGTAAACAGATGCGTGGCGCAGGCACACACACAGCAGCCAGCTGTG	3122
Db	56331	CCGCGCCGGGTCTGTAAACAGATGCGTGGCGCAGGCACACACACAGCAGCCAGCTGTG	56272
QY	3123	TGTGGCGCGGAGTCCGGTGGCGGTCCCAGGTGAGCAGGCGGTGGTGGTGGCGGGCAG	3182
Db	56271	TGTGGCGCGGAGTCCGGTGGCGGTCCCAGGTGAGCAGGCGGTGGTGGTGGCGGGCAG	56212
QY	3183	AGCCATTGTTCCAGCGGTACCGAGCCCCCGCGCTCGCCCGGAGGAGGCGGGGCTTC	3242
Db	56211	AGCCATTGTTCCAGCGGTACCGAGCCCCCGCGCTCGCCCGGAGGAGGCGGGGCTTC	56152
QY	3243	CCGCGT-CCCCAAGTCCAGATCTCTGGGTGGCTCCACGTCTCCCTGCCAGCGCCTGG	3301
Db	56151	CCGCGTCCGCCAAGTCCAGATCTCTGGGTGGCTCCACGTCTCCCTGCCAGCGCCTGG	56092
QY	3302	GGGACGGGAACACGGGACGGAGATGTAGTGGTGGGCGCCCCCGAGGGGTTCACCATG	3361
Db	56091	GGGACGGGAACACGGGACGGAGATGTAGTGGTGGGCGCCCCCGAGGGGTTCACCATG	56032
QY	3362	TTTCTTGAGAAATTTCCCAAGTGCACCCACCCGTTCTCGGTGTGCCGAGGGCGCGTC	3421
Db	56031	TTTCTTGAGAAATTTCCCAAGTGCACCCACCCGTTCTCGGTGTGCCGAGGGCGCGTC	55972
QY	3422	CTGGGCTAGGCTCCGCGGCCACCCAGCCCAACCGGGTCCCAAGCCCTTCCAGAGAGAAAG	3481
Db	55971	CTGGGCTAGGCTCCGCGGCCACCCAGCCCAACCGGGTCCCAAGCCCTTCCAGAGAGAAAG	55912
QY	3482	CTCCCGACGCGGATTCGCGGGCAGAGGCCACAGCGCGGGTGAAGAGAAAGCTGAGAAGGA	3541
Db	55911	CTCCCGACGCGGATTCGCGGGCAGAGGCCACAGCGCGGGTGAAGAGAAAGCTGAGAAGGA	55852
QY	3542	GAACAGAGGGAGGGGAGCGAGAGTGTGCGGAGAGGGAACAGCAGATTGCGCCGAG	3601
Db	55851	GAACAGAGGGAGGGGAGCGAGAGTGTGCGGAGAGGGAACAGCAGATTGCGCCGAG	55792
QY	3602	CCAAATGGCAACGGCAGGACGAGGTGGCACCAAAATTCCTTCGGCCATGACGAGCCGGAG	3661
Db	55791	CCAAATGGCAACGGCAGGACGAGGTGGCACCAAAATTCCTTCGGCCATGACGAGCCGGAG	55732
QY	3662	TTTACAGAAAGCCTCATTAGCATTTCCCAAGGACAGGGGACAGGGCCCGGGTGGT	3721
Db	55731	TTTACAGAAAGCCTCATTAGCATTTCCCAAGGACAGGGGACAGGGCCCGGGTGGT	55672
QY	3722	GTGGTGTGCGGTGTGCGGACATCCCGGCGCCCTGCTGCTGGTTCGCGGAGCCTTCGCGCT	3781
Db	55671	GTGGTGTGCGGTGTGCGGACATCCCGGCGCCCTGCTGCTGGTTCGCGGAGCCTTCGCGCT	55612
QY	3782	CTGTCTCTCCCTCCCTCCCGCCCTTACCTCCACGCGGACCGCCCGCCAGTCAACTCCT	3841
Db	55611	CTGTCTCTCCCTCCCTCCCGCCCTTACCTCCACGCGGACCGCCCGCCAGTCAACTCCT	55552
QY	3842	CGCACTTTGCCCCCTGCTTGGCAGCGGATAAAGGGGGCTGAGGAATAACCGGACACGGTC	3901
Db	55551	CGCACTTTGCCCCCTGCTTGGCAGCGGATAAAGGGGGCTGAGGAATAACCGGACACGGTC	55492
QY	3902	ACCGTGTGCCAGCTTAGCCCTTTAAATTCGCGGTCCGGGACCTCCACGACCGCGGCTA	3961
Db	55491	ACCGTGTGCCAGCTTAGCCCTTTAAATTCGCGGTCCGGGACCTCCACGACCGCGGCTA	55432
QY	3962	CGCGCGCAACACCACTAGCGGTGAAGCGCGCGGGCTCAGCGGTACCGCGGGGCTTCGA	4021

Db 55431 GC CGC GAC ACCACTAGCTGCTCAAGCGCCGCGGCTCAGCGCGGCTTCCA 55372

QY 4022 AACCGCAGTCTCCGGGACACCCGAACTCCGCTCCGGAGCCTCAGCCCTCGAAAGTGA 4081

Db 55371 AACCGCAGTCTCCGGGACACCCGAACTCCGCTCCGGAGCCTCAGCCCTCGAAAGTGA 55312

QY 4082 TCCCGGCATCCGAGACCAAGATCCGGCCACATTGCTGCAGGACGATGTCAGTTTCCCA 4141

Db 55311 TCCCGGCATCCGAGACCAAGATCCGGCCACATTGCTGCAGGACGATGTCAGTTTCCCA 55252

QY 4142 GCTGTGCC 4150

Db 55251 GCTGTGCC 55243

RESULT 3

AL139819

LOCUS

DEFINITION Human DNA sequence from clone RP11-34D15 on chromosome 10, complete sequence.

ACCESSION AL139819

VERSION AL139819.8

KEYWORDS GI:15705162

SOURCE HTP.

ORGANISM human.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 115278)

AUTHORS Tracey, A.

TITLE Direct Submission

JOURNAL Submitted (18-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT requests: clonerequest@sanger.ac.uk

On Sep 19, 2001 this sequence version replaced gi:14586033.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormep

This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr10

RP11-34D15 is from the library RPCI1-11.1 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-34D15 it may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true right end of clone RP11-34D15 is at 115278 in this sequence. The true left end of clone RP11-285F16 is at 56886 in this sequence. The true right end of clone RP11-316M21 is at 2000 in this sequence.

FEATURES

Location/Qualifiers

1..115278

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="10"

/clone="RP11-34D15"

|||||  
Db 65552 GAGCAAGGCTCTACTATGTTGGCATGATCCGCCACCTCCACTCCCAAGTGCATC 65611  
QY 1023 TTAICTGTTCAATAGTCAGTTGACAGCAATTTAGGTGTTTCCACTTTTGGACCAATTAIG 1082  
Db 65612 TTATCTGTTCATTAAGTCAGTTGACAGCAATTTAGGTGTTTCCACTTTTGGACCAATTAIG 65671  
QY 1083 AATAATACCTCAGTGAATTAATCATGTATACATTTGTTGGGCATATGTTTTCATTTCTGT 1142  
Db 65672 AATAATACCTCAGTGAATTAATCATGTATACATTTGTTGGGCATATGTTTTCATTTCTGT 65731  
QY 1143 TGGGTTTATATCTAGGAGTGAATTTGTTGGATCCCGGTATATTTTACAGAGCAGAGTT 1202  
Db 65732 TGGGTTTATATCTAGGAGTGAATTTGTTGGATCCCGGTATATTTTACAGAGCAGAGTT 65791  
QY 1203 CAGGGGAGAGAAACTTGGGAAATGAAGCAATGTTTGAATAATCAGCAAGAGTGCAGGGGT 1262  
Db 65792 CAGGGGAGAGAAACTTGGGAAATGAAGCAATGTTTGAATAATCAGCAAGAGTGCAGGGGT 65851  
QY 1263 TTTTCGGAGTTTATTTTATATCTGTTGACAAATGTCAGTTTGATCAAGATCAAGTT 1322  
Db 65852 TTTTCGGAGTTTATTTTATATCTGTTGACAAATGTCAGTTTGATCAAGATCAAGTT 65911  
QY 1323 ATACTAGTGAGAGTGAATTAAGGCTGGAATAGGCGCTTCAGAGTAAATCATGAAG 1382  
Db 65912 ATACTAGTGAGAGTGAATTAAGGCTGGAATAGGCGCTTCAGAGTAAATCATGAAG 65971  
QY 1383 CACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAACAAAAATTAATAAAATCAAAAT 1442  
Db 65972 CACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAACAAAAATTAATAAAATCAAAAT 66031  
QY 1443 TTTTTTTTTTTTTGAGAAAGCTTTCCTTTTCACCCCTGGCTGGAGGCGAGTGGTG 1502  
Db 66032 TTTTTTTTTTTTTGAGAAAGCTTTCCTTTTCACCCCTGGCTGGAGGCGAGTGGTG 66091  
QY 1503 ATCTCAGCTCACTGCACTTTCGCTCCCGGTTTCAGCAATTTCTCGTTTCAGCCCTC 1562  
Db 66092 ATCTCAGCTCACTGCACTTTCGCTCCCGGTTTCAGCAATTTCTCGTTTCAGCCCTC 66151  
QY 1563 CAAGTAGCTGGACTACAGCACTTCCACCATGCCAGCTGATTTTGTATTTTATGTA 1622  
Db 66152 CAAGTAGCTGGACTACAGCACTTCCACCATGCCAGCTGATTTTGTATTTTATGTA 66211  
QY 1623 GAGATGGGATTTCACTTTGTTGGCAAGCTGGTCTCAAACTTTTGTGTCATAATTTGTT 1682  
Db 66212 GAGATGGGATTTCACTTTGTTGGCAAGCTGGTCTCAAACTTTTGTGTCATAATTTGTT 66271  
QY 1683 GTAACTATTCTTCTTTGCTGAGGTAGGCGCCCGACCAAAATAAATCTTAGAA 1742  
Db 66272 GTAACTATTCTTCTTTGCTGAGGTAGGCGCCCGACCAAAATAAATCTTAGAA 66331  
QY 1743 TCCAAATCAGTGTGTTGGTTTGACCACTGTCACTTTGAGAACCAAGTGTGACCAAGGSCCT 1802  
Db 66332 TCCAAATCAGTGTGTTGGTTTGACCACTGTCACTTTGAGAACCAAGTGTGACCAAGGSCCT 66391  
QY 1803 CAGAGTAGAGGTGATCTGCTCGAAAGAGAAATAGAATGAAATAATTTCCGGGCCAG 1862  
Db 66392 CAGAGTAGAGGTGATCTGCTCGAAAGAGAAATAGAATGAAATAATTTCCGGGCCAG 66451  
QY 1863 GCGTGTGGCTCATGCTGCTTAATCCAGCACTTTGGAGGCCAAGGATGGATCACCT 1922  
Db 66452 GCGTGTGGCTCATGCTGCTTAATCCAGCACTTTGGAGGCCAAGGATGGATCACCT 66511  
QY 1923 GAGTCAGAGGTTCAAAACAGGCTGGCCCAACATGTTGTAACCCCGTCTCTACTAAAAAT 1982  
Db 66512 GAGTCAGAGGTTCAAAACAGGCTGGCCCAACATGTTGTAACCCCGTCTCTACTAAAAAT 66571  
QY 1983 ACAAAAATTAGTGAAGTGGTGGGCGCATGCTGTAATCCCGTACTTTGGAGGGTGA 2042  
Db 66572 ACAAAAATTAGTGAAGTGGTGGGCGCATGCTGTAATCCCGTACTTTGGAGGGTGA 66631  
QY 2043 GGCAGGAGAAATTTCTGAACCCCGGAGGAGAGTTGCAGTGAAGGAGATCACCACT 2102  
|||||

Db 66632 GGCAGGAGAAATTTCTGTGAACCCCGGAGGCGAGGTTGCAGTGAAGCGAGATCACACCACT 66691  
QY 2103 GCACTCCAGCCTGGGGAGAGAGAGACTTCCCTCTCAAAAAACAAAAACAAAGAAAT 2162  
Db 66692 GCACTCCAGCCTGGGGAGAGAGAGACTTCCCTCTCAAAAAACAAAAACAAAGAAAT 66751  
QY 2163 TAAGCAAAATTAGACATTTGCAGAGAGACCTGAAGGGGTCAGACCACTACAGATTTCTG 2222  
Db 66752 TAAGCAAAATTAGACATTTGCAGAGAGACCTGAAGGGGTCAGACCACTACAGATTTCTG 66811  
QY 2223 TGCCACATGCCAACTACTTCTGAGGCATGACTGGATGAGCTGCCATCTGAAATCATC 2282  
Db 66812 TGCCACATGCCAACTACTTCTGAGGCATGACTGGATGAGCTGCCATCTGAAATCATC 66871  
QY 2283 CAGCTTTGTTTCAGAACTTTTCACACGGACAGGAGCCAGGACTGGAATGCACTCTCTGG 2342  
Db 66872 CAGCTTTGTTTCAGAACTTTTCACACGGACAGGAGCCAGGACTGGAATGCACTCTCTGG 66931  
QY 2343 TCAGTGGCCAGAGAGTTGGCCTTTCACCTTGAGACCACTGAGCCCAACAAAGGAGCTGCTTAG 2402  
Db 66932 TCAGTGGCCAGAGAGTTGGCCTTTCACCTTGAGACCACTGAGCCCAACAAAGGAGCTGCTTAG 66990  
QY 2403 TCTACTCTCCAGGAAATCCCAAGTGTCTCTCTGGGAAGTGAATCATTTGGCGCAGCA 2462  
Db 66991 TCTACTCTCCAGGAAATCCCAAGTGTCTCTCTGGGAAGTGAATCATTTGGCGCAGCA 67050  
QY 2463 CTCGTAATTTCTCTCTTCCAGGGGAAGATCTTAGGCGAGTATTTGGGAAAGACATG 67110  
Db 67051 CTCGTAATTTCTCTCTTCCAGGGGAAGATCTTAGGCGAGTATTTGGGAAAGACATG 67170  
QY 2523 GGCATGAAGGACACCGGCTGAATGCATAGCCTGCCTGCTGCTGCTCATCTCATGTAAG 2582  
Db 67171 GGCATGAAGGACACCGGCTGAATGCATAGCCTGCCTGCTGCTGCTCATCTCATGTAAG 67170  
QY 2583 CTCTACAGACACCGGAAAGATGGGGCACAGGACAGATCAGTAGGCTCAGAGCACTCTC 2642  
Db 67171 CTCTACAGACACCGGAAAGATGGGGCACAGGACAGATCAGTAGGCTCAGAGCACTCTC 67230  
QY 2643 AGGACCGAGGGCAATATGTCCTGAGCAGGAGTAAGAGCTTGGGCTCTCATATGTTGT 2702  
Db 67231 AGGACCGAGGGCAATATGTCCTGAGCAGGAGTAAGAGCTTGGGCTCTCATATGTTGT 67290  
QY 2703 TTCTGGGCTCAACTGCCAGCTCCGCTCACTTACTGTTGCTGTGACCATGGGCAAGTTAT 2762  
Db 67291 TTCTGGGCTCAACTGCCAGCTCCGCTCACTTACTGTTGCTGTGACCATGGGCAAGTTAT 67350  
QY 2763 CCATCTCTCATATCTCTTCTCTCACTTTTAAATGGAATTAATGGGTACCCACTCCCA 2822  
Db 67351 CCATCTCTCATATCTCTTCTCTCACTTTTAAATGGAATTAATGGGTACCCACTCCCA 67410  
QY 2823 GGTTCACAGAGGCTTACAGAAACGATTTCTGTGAATTTGGCTTGCAGTAATTAATCAA 2882  
Db 67411 GGTTCACAGAGGCTTACAGAAACGATTTCTGTGAATTTGGCTTGCAGTAATTAATCAA 67470  
QY 2883 TACTTGCCAGTATTTCTTATTTCCACATCCAGGCCCTTTCGCTGCTGCTGGGTGAAACA 2942  
Db 67471 TACTTGCCAGTATTTCTTATTTCCACATCCAGGCCCTTTCGCTGCTGCTGGGTGAAACA 67530  
QY 2943 CATGTAGTGTTCCTGAGCAGGTTTCCAAAGAGATTCAAAATTTACACCTGCCAGTC 3002  
Db 67531 CATGTAGTGTTCCTGACGCTTCCAAAGAGATTCAAAATTTACACCTGCCAGTC 67590  
QY 3003 TGAAGATCTCCAAAACATCCCGCAGCATCTTGGAGGCGCGGCTTGGGATGGGACTG 3062  
Db 67591 TGAAGATCTCCAAAACATCCCGCAGCATCTTGGAGGCGCGGCTTGGGATGGGACTG 67650  
QY 3063 CCCGCCCGGCTCTGAACGATTCGTCGCGCAGGACACACACACACAGCCAGCCTGTG 3122  
Db 67651 CCCGCCCGGCTCTGAACGATTCGTCGCGCAGGACACACACACACAGCCAGCCTGTG 67710  
QY 3123 TGTGCGCGGAGTCCGGTCCGGTCCGGTGAAGCGCGGTGGCTGGTGGGCGGGGCGAG 3182  
Db 67711 TGTGCGCGGAGTCCGGTCCGGTCCGGTGAAGCGCGGTGGCTGGTGGGCGGGGCGAG 67770

```
QY 3183 AGCAATGTTCCAGCGGTACGAGCCCGCCCGCTCCCGCGGAGGAGCGGGGCTTC 3242
|||||
Db 67771 AGCAATGTTCCAGCGGTACGAGTCCCGCCCGCTCCCGCGGAGGAGCGGGGCTTC 67830
|||||
QY 3243 CGCGCT-CCCCAAGCTCCAGATCTTGGGTGGCTGCTCCAGCTCTCCCTGCACGCGCGCTGG 3301
|||||
Db 67831 CGCGCTCCCCAAGCTCCAGATCTTGGGTGGCTGCTCCAGCTCTCCCTGCACGCGCGCTGG 67890
|||||
QY 3302 GGGGACGGGAAGACGGGACGGAGATGTTAGTGGTGGGCGCCCGCGAGGTTTCCACTG 3361
|||||
Db 67891 GGGGAC-GGAAGACGGGACGGAGATGTTAGTGGTGGGCGCCCGCGAGGTTTCCACTG 67949
|||||
QY 3362 TTCTCTGAGAACTTCCCGAGTGGCCACCCACCCGCTTCTCGTGGCCGAGGCGGCTG 3421
|||||
Db 67950 TTCTCTGAGAACTTCCCGAGTGGCCACCCACCCGCTTCTCGTGGCCGAGGCGGCTG 68009
|||||
QY 3422 CTGGCTAGGCTCCGCGCCCGCCAGCCCAAAACCGGGTCCCGAGCCCTTCCAGAGAGAA 3481
|||||
Db 68010 CTGGCTAGGCTCCGCGCCCGCCAGCCCAAAACCGGGTCCCGAGCCCTTCCAGAGAGAA 68069
|||||
QY 3482 CTCCGACGCGGGATGCGCGGAGAGGCCAGCGGGTGGAGAGAGCTGAGAAGA 3541
|||||
Db 68070 CTCCGACGCGGGATGCGCGGAGAGGCCAGCGGGTGGAGAGAGCTGAGAAGA 68129
|||||
QY 3542 GAAACAGAGGGGAGGGGAGCGAGAGCTGGCGGAGAGGGAAACAGAGATTGCGCCGAG 3601
|||||
Db 68130 GAAACAGAGGGGAGGGGAGCGAGAGCTGGCGGAGAGGGAAACAGAGATTGCGCCGAG 68189
|||||
QY 3602 CCAATGCAACGGCAGGAGCGAGGTGGCCACCAATTTCCCTTCCGCGCAATGACGAGCGGAG 3661
|||||
Db 68190 CCAATGCAACGGCAGGAGCGAGGTGGCCACCAATTTCCCTTCCGCGCAATGACGAGCGGAG 68249
|||||
QY 3662 TTTACAGAGCTCATTTAGCATTTCCCGAGGAGCGAGGGGAGCGAGGCGGGTGGT 3721
|||||
Db 68250 TTTACAGAGCTCATTTAGCATTTCCCGAGGAGCGAGGGGAGCGAGGCGGGTGGT 68309
|||||
QY 3722 GTGGTGTGGGTGCGGAGCATCCCGCGCCCTGCTGCGGTCGCGGAGCGCTCGGCT 3781
|||||
Db 68310 GTGGTGTGGGTGCGGAGCATCCCGCGCCCTGCTGCGGTCGCGGAGCGCTCGGCT 68369
|||||
QY 3782 GTGTCTCTCCCGCTCCCGCGCTTACCTCCAGCGGGGACCGCGCCGCGCACTCACTCT 3841
|||||
Db 68370 GTGTCTCTCCCGCTCCCGCGCTTACCTCCAGCGGGGACCGCGCCGCGCACTCACTCT 68429
|||||
QY 3842 CGCACTTTCGCGCTGTTGGCAGCGATTAAGAGGGGCTGAGGAATATCCGAGCAGGTC 3901
|||||
Db 68430 CGCACTTTCGCGCTGTTGGCAGCGATTAAGAGGGGCTGAGGAATATCCGAGCAGGTC 68489
|||||
QY 3902 ACCGCTTCCAGCTAGCTAGCTTTAAATTTCCCGGCTCGGGGACCTCCAGCAGCGCGGCTA 3961
|||||
Db 68490 ACCGCTTCCAGCTAGCTAGCTTTAAATTTCCCGGCTCGGGGACCTCCAGCAGCGCGGCTA 68549
|||||
QY 3962 GCGCGGACAACTAGCTAGCTGTCAGAGCGCGCGGCTCAGCGCGTACCGGGGGCTTCCA 4021
|||||
Db 68550 GCGCGGACAACTAGCTAGCTGTCAGAGCGCGCGGCTCAGCGCGTACCGGGGGCTTCCA 68609
|||||
QY 4022 AACCGCTTCTCCGCGACCGCACTTCCGCTCCGAGGCTCAGCCCGCTTGGAAAGTGA 4081
|||||
Db 68610 AACCGCTTCTCCGCGACCGCACTTCCGCTCCGAGGCTCAGCCCGCTTGGAAAGTGA 68669
|||||
QY 4082 TCCCGGATCGGAGACCAAGATGCGGCGCACTTCTGCTGAGGAGCATGTGAGTTTCCA 4141
|||||
Db 68670 TCCCGGATCGGAGACCAAGATGCGGCGCACTTCTGCTGAGGAGCATGTGAGTTTCCA 68729
|||||
QY 4142 GCTTGGCCC 4150
|||||
Db 68730 GCTTGGCCC 68738
|||||
```

RESULT 4  
AL391723/c  
LOCUS AL391723 185644 bp DNA linear HTG 11-SEP-2001

DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Homo sapiens chromosome 10 clone RP11-3B21, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 14 unordered pieces.  
AL391723  
AL391723.7 GI:10186824  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Barton, J.  
Direct Submission  
Submitted (10-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Sep 19, 2000 this sequence version replaced gi:10178830.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BA38B21  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 175002 bases at least Q40  
Consensus quality: 179431 bases at least Q30  
Consensus quality: 182156 bases at least Q20  
Insert size: 184344; sum-of-contigs  
Insert size: 140788; 6.0% error; agarose-fp  
Quality coverage: 4.15x in Q20 bases; sum-of-contigs Quality  
coverage: 6.51x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 59110: contig of 59110 bp in length  
\* 59111 59210: gap of 100 bp  
\* 59211 61285: contig of 2075 bp in length  
\* 61286 61385: gap of 100 bp  
\* 61386 64379: contig of 2994 bp in length  
\* 64380 64479: gap of 100 bp  
\* 64480 66747: contig of 2268 bp in length  
\* 66748 66847: gap of 100 bp  
\* 66848 114355: contig of 47508 bp in length  
\* 114356 114455: gap of 100 bp  
\* 114456 125406: contig of 10951 bp in length  
\* 125407 125506: gap of 100 bp  
\* 125507 127673: contig of 2167 bp in length  
\* 127674 127773: gap of 100 bp  
\* 127774 130111: contig of 2338 bp in length  
\* 130112 130211: gap of 100 bp  
\* 130212 132354: contig of 2143 bp in length  
\* 132355 132454: gap of 100 bp  
\* 132455 136537: contig of 4083 bp in length  
\* 136538 136637: gap of 100 bp  
\* 136638 139127: contig of 2490 bp in length  
\* 139128 139227: gap of 100 bp  
\* 139228 181355: contig of 42128 bp in length  
\* 181356 181455: gap of 100 bp  
\* 181456 183524: contig of 2069 bp in length  
\* 183525 183624: gap of 100 bp  
\* 183625 185644: contig of 2020 bp in length.  
Location/Qualifiers  
1. 185644  
/organism="Homo sapiens"



QY	3457	CTTCCAGAGAGAAAGCTCCCGACCGGGATGCCGGGCAGAGGCCCGGGGGTGGGAAG	3526
Db	182146	CTTCCAGAGAGAAAGCTCCCGACCGGGATGCCGGGCAGAGGCCCGGGGGTGGGAAG	182087
QY	3527	AGAAGCTCAGAGAGAGAAACAGAGGGGAGGGGACGAGGAGCTGGCGCAGAGGGGAACA	3586
Db	182086	AGAAGCTCAGAGAGAGAAACAGAGGGGAGGGGACGAGGAGCTGGCGCAGAGGGGAACA	182027
QY	3587	GCAGATTCCGCCCGAGCCCAATGGCAACGCGACGAGCAGAGTGGCACCAATTCCTTCGCGCC	3646
Db	182026	GCAGATTCCGCCCGAGCCCAATGGCAACGCGACGAGCAGAGTGGCACCAATTCCTTCGCGCC	181967
QY	3647	AATCAGCAGSCGGAGTTTACAGAGCCCTCATTCAGATTTCCTCCAGAGCAGGGCAGGGG	3706
Db	181966	AATCAGCAGSCGGAGTTTACAGAGCCCTCATTCAGATTTCCTCCAGAGCAGGGCAGGGG	181907
QY	3707	CAGAGCCCGGTGTGTGTGTCTCGGTGTCTGCGCAGCATCCCGCGGCCCTGCTCGGGTCTGC	3766
Db	181906	CAGAGCCCGGTGTGTGTGTCTCGGTGTCTGCGCAGCATCCCGCGGCCCTGCTCGGGTCTGC	181847
QY	3767	CGCAGAGCTCGGCCTCTGTCTCTCCCTCTCCGCGCCTTAAGTCCACGCGGACCGGCCG	3826
Db	181846	CGCAGAGCTCGGCCTCTGTCTCTCCCTCTCCGCGCCTTAAGTCCACGCGGACCGGCCG	181787
QY	3827	CGCCAGTCAACTCTCGCACATTTCGCCCTGCTTGGCAGCGGATAAAGGGGGCTCAGGAA	3886
Db	181786	CGCCAGTCAACTCTCGCACATTTCGCCCTGCTTGGCAGCGGATAAAGGGGGCTCAGGAA	181727
QY	3887	ATACCGGACAGGCTACCCGTTGCCAGCTCTAGCCTTTAAATTCCTCGGCTCGGGGACCTC	3946
Db	181726	ATACCGGACAGGCTACCCGTTGCCAGCTCTAGCCTTTAAATTCCTCGGCTCGGGGACCTC	181667
QY	3947	CAGCACCGGGCTAGCGCGACACACAGCTAGCTGCAAGGCGCGGGCTCAGCGGT	4006
Db	181666	CAGCACCGGGCTAGCGCGCGACACACAGCTAGCTGCAAGGCGCGGGCTCAGCGGT	181607
QY	4007	ACCGCGGGCTTCGAAACCGCAGTCTCTCCGCGCACCCCGAACTCCGGTCCCGAGGCTCAG	4066
Db	181606	ACCTCGGGCTTCGAAACCGCAGTCTCTCCGCGCACCCCGAACTCCGGTCCCGAGGCTCAG	181547
QY	4067	CCCCCTGGAAGTATCCCGGCATCCGAGACCAAGATGCCGGCCACTTCTCTCGCAGAC	4126
Db	181546	CCCCCTGGAAGTATCCCGGCATCCGAGACCAAGATGCCGGCCACTTCTCTCGCAGAC	181487
QY	4127	GATGTGAGTTTCCAGCGCTGGCCC	4150
Db	181486	GAGTGAGTTTCCAGCGCTGGCCC	181463
RESULT	5		
LOCUS	AX431388	1970 bp	DNA
DEFINITION	Sequence 21 from Patent WO0240666.	linear	PAT 28-JUN-2002
ACCESSION	AX431388		
VERSION	AX431388.1	GI:21656235	
KEYWORDS	human.		
SOURCE	homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1		
AUTHORS	Enkins,D.K., Winther,M.D., Haardt,M., Goldberg,Y.P., Nwaka,S.O., Ponton,A., Allen,S.J., de Antueno,R.J. and Knickle,L.C.		
TITLE	Fat regulated genes, uses thereof, and compounds for modulating same		
JOURNAL	Patent: WO 0240666-A 21 23-MAY-2002;		
FEATURES	XENON GENETICS INC (CA)		
source	Location/Qualifiers		
	l..1970		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	428 a 590 c 588 g 364 t		
ORIGIN			







```

* 12254 14386: contig of 2133 bp in length
* 14387 14486: gap of unknown length
* 14487 16990: contig of 2504 bp in length
* 16991 17090: gap of unknown length
* 17091 19022: contig of 1932 bp in length
* 19023 19122: gap of unknown length
* 19123 20497: contig of 1375 bp in length
* 20498 20597: gap of unknown length
* 20598 22209: contig of 1612 bp in length
* 22210 22309: gap of unknown length
* 22310 24677: contig of 2368 bp in length
* 24678 24777: gap of unknown length
* 24778 27656: contig of 2878 bp in length
* 27657 27756: gap of unknown length
* 27757 30189: contig of 2434 bp in length
* 30190 30289: gap of unknown length
* 30290 33658: contig of 3369 bp in length
* 33659 33758: gap of unknown length
* 33759 36912: contig of 3154 bp in length
* 36913 37011: gap of unknown length
* 37012 41069: contig of 4057 bp in length
* 41070 41169: gap of unknown length
* 41170 43658: contig of 2489 bp in length
* 43659 46574: contig of 2816 bp in length
* 46575 46674: gap of unknown length
* 46675 49890: contig of 3216 bp in length
* 49891 49990: gap of unknown length
* 49991 53751: contig of 3761 bp in length
* 53752 53851: gap of unknown length
* 53852 57378: contig of 3527 bp in length
* 57379 57478: gap of unknown length
* 57479 61598: contig of 4120 bp in length
* 61599 61698: gap of unknown length
* 61699 65686: contig of 3988 bp in length
* 65687 65786: gap of unknown length
* 65787 70673: contig of 4887 bp in length
* 70674 70773: gap of unknown length
* 70774 76443: contig of 5669 bp in length
* 76443 76542: gap of unknown length
* 76543 80193: contig of 3651 bp in length
* 80194 80293: gap of unknown length
* 80294 83630: contig of 3337 bp in length
* 83631 83730: gap of unknown length
* 83731 89197: contig of 5466 bp in length
* 89198 89296: gap of unknown length
* 89297 93491: contig of 4195 bp in length
* 93492 93591: gap of unknown length
* 93592 99552: contig of 6361 bp in length
* 99553 100052: gap of unknown length
* 100053 107237: contig of 7185 bp in length
* 107238 107337: gap of unknown length
* 107338 113752: contig of 6415 bp in length
* 113753 113852: gap of unknown length
* 113853 119355: contig of 5503 bp in length
* 119356 119455: gap of unknown length
* 119456 127315: contig of 7860 bp in length
* 127316 127415: gap of unknown length
* 127416 137358: contig of 9943 bp in length
* 137359 137458: gap of unknown length
* 137459 148311: contig of 10853 bp in length
* 148312 148411: gap of unknown length
* 148412 160062: contig of 11651 bp in length
* 160063 160162: gap of unknown length
* 160163 179879: contig of 19717 bp in length.
FEATURES
    source
    1. .179879
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="3"
    /map="3p"
    /clone="RP11-799L19"
    1. .1496
misc_feature

```

```

misc_feature /note="assembly_name:Contig7"
1597. .2949
/note="assembly_name:Contig9"
3050. .4409
/note="assembly_name:Contig10"
4510. .6137
/note="assembly_name:Contig11"
6238. .7284
/note="assembly_name:Contig12
clone_end.SP6
vector_side:right"
7385. .9013
/note="assembly_name:Contig13"
9114. .10444
/note="assembly_name:Contig14"
10545. .12153
/note="assembly_name:Contig15"
12254. .14386
/note="assembly_name:Contig16"
14487. .16990
/note="assembly_name:Contig17"
17091. .19022
/note="assembly_name:Contig18"
19123. .20497
/note="assembly_name:Contig19"
20598. .22209
/note="assembly_name:Contig20"
22310. .24677
/note="assembly_name:Contig21"
24778. .27655
/note="assembly_name:Contig22"
27756. .30189
/note="assembly_name:Contig23"
30290. .33658
/note="assembly_name:Contig24"
33759. .36912
/note="assembly_name:Contig25"
37013. .41069
/note="assembly_name:Contig26"
41170. .43658
/note="assembly_name:Contig27"
43759. .46574
/note="assembly_name:Contig28"
46675. .49890
/note="assembly_name:Contig29"
49991. .53751
/note="assembly_name:Contig30"
53852. .57378

```

Query Match 34.9%; Score 1448.2; DB 2; Length 179879;  
 Best Local Similarity 99.1%; Pred. No. 0;  
 Matches 1456; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

QY 3 GAATTCATCCATTAAATCATACAAATTAATGGCTTTTAGTATATATTCACAGTTGTCAT 62
|||||
Db 109955 GAATTCATCCATTAAATCATACAAATTAATGGCTTTTAGTATATATTCACAGTTGTCAT 110014
QY 53 CCATCACAAATCCATTTTAGAACAGTTTATTACTCCAAAATAAACCCTGCATTCCTTAG 122
|||||
Db 110015 CCATCACAAATCCATTTTAGAACAGTTTATTACTCCAAAATAAACCCTGCATTCCTTAG 110074
QY 123 CCATCACGCCCAACATCCTCCATCCTCTCCAAAGCCCTGGCAACCAATCTACTT 182
|||||
Db 110075 CCATCACGCCCAACATCCTCCATCCTCTCTCCAAAGCCCTGGCAACCAATCTACTT 110134
QY 133 TCTGTCTCTATAAATTTGCCAATTTCCGACATTTTCATATTAATGAAGCAACCAATCT 242
|||||
Db 110135 TCTGTCTCTATAAATTTGCCAATTTCCGACATTTTCATATTAATGAAGCAACCAATCT 110194
QY 243 GAGACTTTGTGACGTGGCTGCTTTCACATTTAGCATTTCTATTTTAAAGGCTCATTTATGTACA 302
|||||
Db 110195 GAGACTTTGTGACGTGGCTGCTTTCACATTTAGCATTTCTATTTTAAAGGCTCATTTATGTACA 110254

```

QY 303 GTACTTACGACTACTTCTATCTCTTTTATCTCAATGGTATTCACACTGTGTGGGTATCC 362  
|||||  
Db 110255 GTACTTACGACTACTTCTATCTCTTTTATCTCAATGGTATTCACACTGTGTGGGTATCC 110314  
|||||  
QY 363 CATATCATATTTATTTAGAGACAGGTTCTACCTCTGTCACCCAGGTGGAGTGCAGTGGCAC 422  
|||||  
Db 110315 CATATCATATTTATTTAGAGACAGGTTCTACCTCTGTCACCCAGGTGGAGTGCAGTGGCAC 110374  
|||||  
QY 423 AATCATAGCTCAGTGTACCTCAACCTCTGGGCTCAAGTGATCCTACTACTCAGCCTC 482  
|||||  
Db 110375 AATCATAGCTCAGTGTACCTCAACCTCTGGGCTCAAGTGATCCTACTACTCAGCCTC 110434  
|||||  
QY 483 CAGAGTACTAGGACTACAGGCACACAGCATTACCTGGCTGCTATTTTATTTTATTT 542  
|||||  
Db 110435 CAGAGTACTAGGACTACAGGCACACAGCATTACCTGGCTGCTATTTTATTTTATTT 110494  
|||||  
QY 543 TCATTTTATGATTTATTTCTTTCTTTTGTGTTGTTTGTGATAGAGGCTCTCAC 602  
|||||  
Db 110495 TCATTTTATGATTTATTTCTTTCTTTTGTGTTGTTTGTGATAGAGGCTCTCAC 110554  
|||||  
QY 603 TTTCTTACCCAGGCTGGAGGCGAGTGGTCATGGTGCACAGCTGACAGGCTTGACTTCCTGG 662  
|||||  
Db 110555 TTTCTTACCCAGGCTGGAGGCGAGTGGTCATGGTGCACAGCTGACAGGCTTGACTTCCTGG 110614  
|||||  
QY 663 GCTCAAGTGTCTCTCTGCTCAGCTCCAGCTGAGTGGGACTACAAACAGCTGTCCACC 722  
|||||  
Db 110615 GCTCAAGTGTCTCTCTGCTCAGCTCCAGCTGAGTGGGACTACAAACAGCTGTCCACC 110674  
|||||  
QY 723 ATGCTGCTGTATTTTCTTTCTTTGAAACAGGATATCATCTGTGCTCCAGGCTGGAGT 782  
|||||  
Db 110675 ATGCTGCTGTATTTTCTTTCTTTGAAACAGGATATCATCTGTGCTCCAGGCTGGAGT 110734  
|||||  
QY 783 ACAGTGGCTGAATTAATAGCTCACTGAGCTCCCTCTCTGGCTCAAGCATCCGCTGGC 842  
|||||  
Db 110735 ACAGTGGCTGAATTAATAGCTCACTGAGCTCCCTCTCTGGCTCAAGCATCCGCTGGC 110794  
|||||  
QY 843 CTCAGCATCTGTAGTGTGGGACTAGAGGCTGTGCTGACAGGCTGAGTGTAA 902  
|||||  
Db 110795 CTCAGCATCTGTAGTGTGGGACTAGAGGCTGTGCTGACAGGCTGAGTGTAA 110854  
|||||  
QY 903 AATATGATTTTGTATAGAGGAGTCTGCTATGTTGCTCAGGCTGATTTTATTTGTT 962  
|||||  
Db 110855 AATATGATTTTGTATAGAGGAGTCTGCTATGTTGCTCAGGCTGATTTTATTTGTT 110914  
|||||  
QY 963 GAGACAGGCTCTCACTATGTCATCATCCCTCCCTCTGCTGCTGCTGCTGCTGCTGCT 1022  
|||||  
Db 110915 GAGACAGGCTCTCACTATGTCATCATCCCTCCCTCTGCTGCTGCTGCTGCTGCTGCT 110974  
|||||  
QY 1023 TTATCTGTTTATTTAGTGTGACAGATTTAGGTTGTTTCCACATTTTGGACATTTAG 1082  
|||||  
Db 110975 TTATCTGTTTATTTAGTGTGACAGATTTAGGTTGTTTCCACATTTTGGACATTTAG 111034  
|||||  
QY 1083 AATAATCTCCAGTGAATTTATGATTTATGTTGTTGGGCTATGTTTCTATTTCTGT 1142  
|||||  
Db 111035 AATAATCTCCAGTGAATTTATGATTTATGTTGTTGGGCTATGTTTCTATTTCTGT 111094  
|||||  
QY 1143 TGGGTTTATTTAGGAGTGAATTTGCTGATCCCGGTTAATTTTACAGGCGAGGTT 1202  
|||||  
Db 111095 TGGGTTTATTTAGGAGTGAATTTGCTGATCCCGGTTAATTTTACAGGCGAGGTT 111154  
|||||  
QY 1203 CAGGGAAGAAAACCTTGGGAAATGAAGCATGTTTAGAATCAGCAAGAGTCCAGGGGT 1262  
|||||  
Db 111155 CAGGGAAGAAAACCTTGGGAAATGAAGCATGTTTAGAATCAGCAAGAGTCCAGGGGT 111214  
|||||  
QY 1263 TTTTCGAGTTTATTTTATTTTATTTGATGATGTCAGTGTGATGATGATACAGTT 1322  
|||||  
Db 111215 TTTTCGAGTTTATTTTATTTTATTTGATGATGTCAGTGTGATGATGATACAGTT 111274  
|||||  
QY 1323 ATACTAGTGAAGAGTGAAGTAAAGGCTGGAATAGGCGCTTCAGAGTAAATCATGAAG 1382  
|||||  
Db 111275 ATACTAGTGAAGAGTGAAGTAAAGGCTGGAATAGGCGCTTCAGAGTAAATCATGAAG 111334  
|||||  
QY 1383 CACTTTGAATACCAAAATTAAGGAGCTTGCTGTAAACAAAATAAATAAAATCACAATT 1442  
|||||

Db 111335 CACTTTGAATACCAAAATTAAGGAGCTTGCTGTAAACAAAATAAATAAAATCACAATT 111394  
|||||  
QY 1443 TTTTATTTTATTTTGAAGAAGAGTCTTGC 1471  
|||||  
Db 111395 TTTTATTTTAAATAACCGTCATAGTCCAGC 111423  
|||||  
RESULT 7  
AX357255  
LOCUS AX357255 617 bp DNA linear PAT 13-FEB-2002  
DEFINITION Sequence 1 from Patent WO0162954.  
ACCESSION AX357255  
VERSION AX357255.1 GI:18674409  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Brownlie,A.J., Hayden,M.R., Attie,A.D., Ntambi,J.M.,  
Gray-Keller,M.P. and Miyazaki,M.  
TITLE Methods and compositions using stearyl-coa desaturase to identify  
triglyceride reducing therapeutic agents  
JOURNAL Patent: WO 0162954-A 1 30-AUG-2001;  
Xenon Genetics Inc. (CA); WISCONSIN ALUMNI RESEARCH FOUNDATION  
FEATURES (US); University of British Columbia (CA)  
Location/Qualifiers  
source 1. 617  
BASE COUNT 117 a 211 c 205 g 84 t  
ORIGIN  
Query Match 14.8%; Score 612.4; DB 6; Length 617;  
Best Local Similarity 99.8%; Pred. No. 1.5e-139;  
Matches 613; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3459 CCAGGCCCTTCAGAGAGAAAGCTCCGACGCGGGATGCGGGCAGAGGCCAGCGCG 3518  
|||||  
Db 4 CCGCGCCCTTCAGAGAGAAAGCTCCGACGCGGGATGCGGGCAGAGGCCAGCGCG 63  
|||||  
QY 3519 GGTGAAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3578  
|||||  
Db 64 GGTGAAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123  
|||||  
QY 3579 AGGGAACAGAGAGATGCGCGAGCAATGGCAACGCGAGAGAGAGAGAGAGAGAGAG 3638  
|||||  
Db 124 AGGGAACAGAGAGATGCGCGAGCAATGGCAACGCGAGAGAGAGAGAGAGAGAGAG 183  
|||||  
QY 3639 CTTGCGCAATGAGAGCGAGCGAGTTTACAGAGCCTCATAGCATTTCCCGAGGCGAG 3698  
|||||  
Db 184 CTTGCGCAATGAGAGCGAGCGAGTTTACAGAGCCTCATAGCATTTCCCGAGGCGAG 243  
|||||  
QY 3699 GGCAGGCGAGAGCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3758  
|||||  
Db 244 GGCAGGCGAGAGCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 303  
|||||  
QY 3759 GCGGTGCGCGCGAGCTGCGGCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3818  
|||||  
Db 304 GCGGTGCGCGCGAGCTGCGGCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 363  
|||||  
QY 3819 ACCGCGCGCGCGAGTCAACTCTCTGCGACTTTGCGCTGTGCGAGCATCCCGCGCGCTGCT 3878  
|||||  
Db 364 ACCGCGCGCGCGAGTCAACTCTCTGCGACTTTGCGCTGTGCGAGCATCCCGCGCGCTG 423  
|||||  
QY 3879 CTGAGGAATACCGGACACAGCTCACCCTGTCAGAGCTCTAGCTTTAAATTCGCGGCTCG 3938  
|||||  
Db 424 CTGAGGAATACCGGACACAGCTCACCCTGTCAGAGCTCTAGCTTTAAATTCGCGGCTCG 483  
|||||  
QY 3939 GCGACCTCCACGACCGCGGCTAGCGCGGACACACAGCTAGCTAGCTGCAAGCGCGCGGCT 3998  
|||||  
Db 484 GCGACCTCCACGACCGCGGCTAGCGCGGACACACAGCTAGCTAGCTGCAAGCGCGCGGCT 543  
|||||



Db	5965	TTTTTTAGGCTCGT--CAAGTGAAGCAATGAGAGTGGAGAACAAAGCAATCTGTAA	6022
Qy	1371	AAAATCATGAAGCACITTTGAATACCAAAATTAAGGAGCTTGCCTGTAAACAAAATAATA	1430
Db	6023	CTAGTTGTATCAATTAATGTTAAACACAATGTGTTTTTCATTTTAGTGTGCCATTTAG	6082
Qy	1431	AAAATCAAAATTTTTTTTTTTTTTTTGTAGAAGAGTCTGTCTTTCACTCCCTGCCTGGAG	1490
Db	6083	TTTTTTTCTAATTTATTAATTTTTTTGAGATGGAGTCTTGCTCTGT-TGCCAGGCTGGAA	6141
Qy	1491	GGCAGTGTGTGATCTCAGCTCAGCTGCAACTTTCGCTCCCGGGTTTCAAGCAATTTCTCT	1550
Db	6142	TGCATGGCGGATCTCGGCTACATGCAACCTCCAACTTCTGGGTTCAAGTGAATTTCTCT	6201
Qy	1551	GCTTAGCCTCCCAAGTAGCTGGGACTACAGGCACATCCCACTGCGCCAGTCCAGCTGATTTT	1610
Db	6202	GCCTTAGCTCCCAAGTAGCTGGGACTACAGTGCATGCCACCATGCCAGCTAATTTT	6261
Qy	1611	GTATTTTAGTAGAGATGGGATTTACATTTGTTGGCCAAGCTGGTCTCAAACT-----	1663
Db	6262	GTATTTTAGTAAGATGGGTTTCGCCATGTGGCCAGGATGGCTCAAACTCCTGACC	6321
Qy	1664	---TTTGTGCTCATAAATTTGTAACTAATTTGTTCTTTGCTGAGGTAGGGCCCCCAG	1719
Db	6322	TCAGGTATCCACCGCCTTGGCTCCCCAAAGTGTAGGATTAATAGGCGTGAGCCACTGC	6381
Qy	1720	ACCAAAAAAATAAATCTTAGAATCCAATCAGTGTGTTG-----GTTTGACCA	1769
Db	6382	ACCGCGCGCTCACATTTTCCACACATTTTGTGTAGCACATCTCTCCCTAGCACAT	6441
Qy	1769	CTGCACCTTGAGAACCCACAGTGTGACCAGGGCCCTCAGGAGTAGAGGTGATCTGCTCGA	1828
Db	6442	TTTTCTAACCTTATTTCCCAAGAAGAAAGAACACAGCAATCTACAGCCAGTGTCTATT	6501
Qy	1829	AAGAGAAATCAAAATTAATCT-----TTTTTTT-----1853	
Db	6502	GAAAGATGACCTCAAAAGACTTTCTGAACCTAGTGGCATGTGAACAGGTGAGTGTGAGAC	6561
Qy	1854	-----CCGGGCGCAGCGTGGTGGCTGATCGCTGTAATCCCAAGCACTTTGGAGGCGCA	1905
Db	6562	TAAGACAGGTAGACCAGCGTGGTGGCTTAGCGCTGTAATCCAGCACTTTGGAGGCTG	6621
Qy	1906	AGGCATGTGGATCACCCTGAGTTCAGGAGTTCAAAACAGCGCTGGCCAACTGGTGAAC	1965
Db	6622	AGGTAGCCGATCACCCTGAGTTCGGAGTTCAGACCAACCTGACCAAGCTGGAGAAAC	6681
Qy	1966	CCGTCTCTACTAAAAATACAAAAAATTAAGCTAAGTGTGGTGGCCATGCCCTGTAATCCCA	2025
Db	6682	CCGTCTCTACTAAAAATCTA--AATTAGCCAGGTGTGGTGGCCCATGCTGTGAATCCCA	6739
Qy	2026	GCTACTTGGAGGTGAGGCAGGAGAAATTTCTTGA--CCCGGGAGGCAGAGGTTGCAGTG	2084
Db	6740	GCTACTCGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGGCGGAGTTGTGGTG	6799
Qy	2085	AAGCGAGATCACACCTGCTGACCTCAGCGCTGGG--GGAGAGAGCGAGACTTCTCTCAAAA	2143
Db	6800	AGCCAAGATCGGCCATGTAGCTCTACCTTGGGCGAGCAAGAGTGAATCTCCGCTCAAAA	6859
Qy	2144	AACAAAAAACAAGAA	2161
Db	6860	AAAAAAAAAAAAAAAAA	6877

RESULT	9
AC119725	
LOCUS	
DEFINITION	AC119725 183386 bp DNA linear HTG 09-MAY-2002 Homo sapiens chromosome 3 clone RP11-750d18, *** SEQUENCING IN PROGRESS ***, 24 unordered pieces.
ACCESSION	AC119725
VERSION	AC119725.1 GI:20376806
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Homo sapiens.

ORGANISMS

*Homo sapiens*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 183386)

Muzny,D.M., Adams,C., Adio-oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayelle,M., Banks,T., Barbia,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carleton,F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Deun,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojuboan,I., Roife,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taboz,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 183386)

Worley,K.C.

Direct Submission

Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 183386)

Worley,K.C.

Direct Submission

Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: HGIR  
 Center clone name: RP11-759D18  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 169013 bases at least Q40  
 Consensus quality: 174569 bases at least Q30  
 Consensus quality: 178019 bases at least Q20





Db 163996 ACCAGCCTAGCTAATTTTGTATTTTGTAGTAGAGTGGGTTTCTCCATGTTGGCCAGG 164055  
QY 776 CTGGAGTACAGTGGCGTAATAATAGCTCACTGCAGCCTCCCTCCCTGGGCTCAAGCAATC 835  
Db 164056 CTGGTCTCGA-----ACTCCTGACCTCAGGTGGTC 164085  
QY 836 CGCTGGCCTCAGCATCCTCAGTAGCTGGGACTACAGGCTTGTGCC--ACAGCCCCAGCT 893  
Db 164086 CACCCACCTCGACCTCCACAGTCTGGGATTACAGGCTTGTAGCCGGCCGACACCTGGCT 164145  
QY 894 AAGTTTAAATAATGTTTGGTATAGAGAGGCTTCTGCTATGTTGTGCTCAGGCTGTATT 953  
Db 164146 AATTTTAAAGAAATTTTGTAGAGA-CGGTCTGCTTGTGTCACCGAGTTGGAGT 164204  
QY 954 TTTATGTTGAGACA-----AGTCTCACAATGTTGCCATGATCCCCCACTCCAC 1005  
Db 164205 TCAGTTATCCCATCATAGCTCACTGTAAGTGAATCTCTGGCTTAAAGCAATCCTCT 164264  
QY 1006 TTTCCAAAGTGTCACTTATCTGTTTCAATAGTCACTTGACAGACATTTAGTGTGTTCC 1065  
Db 164265 GCGTCACCTTCCAAATGCTGGATACAGGATGAGCCACTGTACCTGGCCAGTTG 164324  
QY 1066 ACTTTTGACCAATTAATGAATAATACCTCAGTGAATATTCATGTATACATTTGTGGGCA 1125  
Db 164325 ACTTCTAGCTCCTCTGTCAGACTCCCGAGCTTACAGTAAAGTCTTCTCACTCT 164384  
QY 1126 TATGTTTTCATTTCTGTTGGTT-TATATCTAGGAGTGAATTCGAGTCCCGGTAAT 1184  
Db 164385 TTGGCTTGGCAAGCAAAAGCTTTCTGTCATGGTCTCTCAGTCTTCTTACCTCAT 164444  
QY 1185 ATTTTGACAGGAGGTTTCAGGGGAAGAAAATCTTGGGAAATGAAGCATGTTTGAAT 1244  
Db 164445 GTTCAGGCAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 164504  
QY 1245 CAGCAAGAGTGCAGGGTTTTTCGGAGTTTTATTTTATCTGTTGACAAATGTGCACT 1304  
Db 164505 CAGATCTTGCCTTAACTCTCATTTCTCTCTGTCATATATATATATATATATATAT 164564  
QY 1305 TTGATGAGATACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1364  
Db 164565 GACTCCTCCTCCTCCTCTTTCACATTTTCTCTATGATACATAATATTTTACATGTT 164624  
QY 1365 CAGAGTAAATCATGAAGCATTGGAATACCAAAATTAAGGAGCT-----TGGCTGTA 1418  
Db 164625 ATGGGTACATGATTGCTACATGATGATGATGATGATGATGATGATGATGATGATG 164684  
QY 1419 ACAAAATTAATAAATACAAATTTTTTTTTTTTTTTTGGAGAAAGAGTCTTCTCTTCA 1478  
Db 164685 CCATTACCTTGATATTTATCTTTTTTTTTTTTTTTTGGAGATGGAGTCTGCTCTGTTG 164744  
QY 1479 CCCTGGCTGGAGGAGTGGTGTGATCTCAGCTCACTGCAACTTTCGCTCCCGGTTCA 1538  
Db 164745 CTCAGCTGGAGTGAATGGCAACATCTCAGTTCACTGCAACTCTGCTCCAGGTTCA 164804  
QY 1539 AGCAATTCCTCCTCCTCAGCCTCCCAAGTAGCTGGGACTTACAGGCACTTCCCACTATGC 1598  
Db 164805 AGTGATTCCTCGCTCAGCCTCCTCAGTAGCTGGATTTACAGGCGTGTGCCACACAC 164864  
QY 1599 CAGCTGATTTTGTATTTTATGAGATGGGATTTCACTTTGTTGGCAAGCTGCTPC 1658  
Db 164865 CGGCTAATTTTGTATTTTATGAGAGGCGGCTTCACTATGTTGGTCAAGCTGCTC 164924  
QY 1659 AAATCTTTGCTGCTATATCTGTTGTAATGTTTCTTTTCTGTTGCTGTTGCTGTTGCT 1714  
Db 164925 AAATCTAATCTAGGCGATCCACCCACTCAGCCTCCCAAAATGCTACAATTTATAGC 164984  
QY 1715 CCAGACCAAAAAAATAATCTTAGAATCCAAATCAGTGTGTTGTTGACCACTGTCA 1774  
Db 164985 GTGAGCACTGACCTGGCCAAATATTTATCATTTCTATGTTTGGAAATATTTCAAGTC 165044  
QY 1775 CTTGAGAACCAAGTGTACAGGCGCTCAGGAGTAGAGGTGATC-----TCTGCTCGAA 1830  
Db 165045 CTTCTCTCTATCTGCTTTGAAATATACTACACATTTCTGCTGGCTATAGTGGCTCCACT 165104

QY 1831 GAGAAATAGAAATGAAATATTC--TCCGGCCAGGCGTGGTGGCTCATGCTGTAATCCCA 1889  
Db 165105 ATCAACATCAATATTTTCTTCTAGGCCAGGCGTGGTGGCTCAGCCTGTAATCCCA 165164  
QY 1890 GCATTTGGAGGCCAAGGCATGTGGATCACCTGAGGTCAAGAGTTCAAAACAGCGCTGG 1949  
Db 165165 GCATTTGGAGGCCAAGGCAGGAGATCACCTGAGTCCGGAGTTTCGAGACCAAGCTGA 165224  
QY 1950 CCAACATGTGAACCCCGCTCTCTACTATAAATACAAAAATTTAGCTAGTGTGGCG 2009  
Db 165225 CCACATGAGAAACCCCGCTCTCTACTATAAATACAAAAATTTAGCCACATGGTAGCA 165284  
QY 2010 CATGCTGTAAATCCAGCTACTTTGGAGGCTGAGGAGGAGGAGAAATTTCTTGAACCCGAG 2069  
Db 165285 CATGCTGTAAATCTCAGCTACTCAGGAGGCTGAGGAGGAGGAGATCGCTTGAACCCGAG 165344  
QY 2070 CGAGAGTTGCAGTGAAGGAGATCACCACTGCCTCCAGCTGGGGA-CAGAGCGA 2128  
Db 165345 CGAGAGTTGCAGTGAAGGAGATCGCGCAATTCGCTCCAGCTGGGCAACAGAGTGA 165404  
QY 2-29 GACTTCTCTCAAAAAACAAAAACAAAAACAAAGAAATTA 2165  
Db 165405 AATTGATCTCAAAAAACAAACACACACAAAAACAGAA 165441

RESULT 11  
AC005037/c

LOCUS

AC005037 Homo sapiens BAC clone RP11-469M7 from 2, complete sequence.

DEFINITION

AC005037

AC005037.2 GI:4827310

HTG.

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 190508)

Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE

99063792

PUBMED

9847074

2 (bases 1 to 190508)

Abbott, A. and Ie, R.

The sequence of Homo sapiens BAC clone RP11-469M7

JOURNAL

Unpublished

3 (bases 1 to 190508)

Waterston, R.H.

Direct Submission

JOURNAL

Submitted (12-JUN-1998) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 190508)

Waterston, R.H.

Direct Submission

JOURNAL

Submitted (14-MAY-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 190508)

Waterston, R.

Direct Submission

JOURNAL

Submitted (30-SEP-2000) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On May 14, 1999 this sequence version replaced gi:3309089.

COMMENT

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)

----- Summary Statistics

Center project name: H\_NH0469M07

-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoe, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACes3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-13J8; the clone sequenced to the right is RP11-91M5. Actual start of this clone is at base position 1 of RP11-469M7; actual end is at 190508 of RP11-469M7.

#### FEATURES

source	Location/Qualifiers
repeat_region	1..190508
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="2"
repeat_region	/map="2"
repeat_region	/clone="RP11-469M7"
repeat_region	/clone_lib="RPCI-11"
repeat_region	155..222
repeat_region	/rpt_family="L2"
repeat_region	982..1299
repeat_region	/rpt_family="Alu"
repeat_region	1502..1533
repeat_region	/rpt_family="(CA)n"
repeat_region	1535..1823
repeat_region	/rpt_family="Alu"
repeat_region	2005..2308
repeat_region	/rpt_family="Alu"
repeat_region	2392..2562
repeat_region	/rpt_family="L1"
repeat_region	3284..3315
repeat_region	/rpt_family="Alu"
repeat_region	3316..3338
repeat_region	/rpt_family="(CAAA)n"
repeat_region	3339..3621
repeat_region	/rpt_family="Alu"
repeat_region	3650..3743
repeat_region	/rpt_family="AT-rich"
repeat_region	4353..4664
repeat_region	/rpt_family="Alu"
repeat_region	4727..5018
repeat_region	/rpt_family="Alu"
repeat_region	5545..5585
repeat_region	/rpt_family="(CATA)n"
repeat_region	5586..5873
repeat_region	/rpt_family="Alu"
repeat_region	5876..6007
repeat_region	/rpt_family="Alu"
repeat_region	6327..6713
repeat_region	/rpt_family="MER21_g"
repeat_region	6924..7294
repeat_region	/rpt_family="MER2_type"
repeat_region	8027..8049
repeat_region	/rpt_family="AT-rich"
repeat_region	8050..8354
repeat_region	/rpt_family="Alu"
repeat_region	8355..8661
repeat_region	/rpt_family="Alu"
repeat_region	8356..8684
repeat_region	/rpt_family="AT-rich"
repeat_region	9450..9752
repeat_region	/rpt_family="Alu"
repeat_region	10371..10680
repeat_region	/rpt_family="Alu"
repeat_region	10762..11038
repeat_region	/rpt_family="Alu"
repeat_region	11039..11109
repeat_region	/rpt_family="Alu"
repeat_region	11138..11410
repeat_region	/rpt_family="Alu"
repeat_region	11446..11663
repeat_region	/rpt_family="L1"
repeat_region	11664..11959
repeat_region	/rpt_family="Alu"
repeat_region	11960..12189
repeat_region	/rpt_family="L1"
repeat_region	12190..12334
repeat_region	/rpt_family="Alu"
repeat_region	12335..12360
repeat_region	/rpt_family="L1"
repeat_region	12464..12514
repeat_region	/rpt_family="L1"
repeat_region	12515..12822
repeat_region	/rpt_family="Alu"
repeat_region	12823..12862
repeat_region	/rpt_family="L1"
repeat_region	12857..13096
repeat_region	/rpt_family="L1"
repeat_region	13101..13330
repeat_region	/rpt_family="MALR"
repeat_region	13331..13634
repeat_region	/rpt_family="Alu"
repeat_region	13635..13692
repeat_region	/rpt_family="MALR"
repeat_region	13704..13990
repeat_region	/rpt_family="L1"
repeat_region	14082..14387
repeat_region	/rpt_family="Alu"
repeat_region	14402..14704
repeat_region	/rpt_family="Alu"
repeat_region	14797..15095
repeat_region	/rpt_family="Alu"
repeat_region	15162..15196
repeat_region	/rpt_family="Mariner"
repeat_region	15606..15681
repeat_region	/rpt_family="L1"
repeat_region	15691..15827
repeat_region	/rpt_family="MER1_type"
repeat_region	15861..16189
repeat_region	/rpt_family="L2"
repeat_region	16426..16446
repeat_region	/rpt_family="(CAAT)n"
repeat_region	16447..16730
repeat_region	/rpt_family="Alu"
repeat_region	17072..17179
repeat_region	/rpt_family="L2"
repeat_region	17212..17504
repeat_region	/rpt_family="Alu"
repeat_region	17869..18139
repeat_region	/rpt_family="Alu"
repeat_region	18168..18464
repeat_region	/rpt_family="Alu"



```

repeat_region 18606..19155
/rpt_family="MaLR"
repeat_region 19454..19767
/rpt_family="Alu"
repeat_region 19807..19949
/rpt_family="MER1_type"
repeat_region 19950..20247
/rpt_family="Alu"
repeat_region 20248..20489

Query Match      10.2%; Score 423.8; DB 9; Length 190508;
Best Local Similarity 60.1%; Pred. No. 8.5e-93;
Matches 969; Conservative 0; Mismatches 547; Indels 97; Gaps 12;

QY 562 TCTTCTCTTTTGTGTTGTTGTTTGTAGATAGGCTCTACCTTTGTTACCCAGGCTGGAG 621
Db 56758 TCTGATTTTGTGTTTGTGTTTGTGAGACAGGTTTCACTCTCTGCCAGGCTGGAG 66699
QY 622 GGCAGTGCATGGTGACAGCTGA--GCAGCCTTGACTTCTCGGGCTCAAGTGAATCCTCCT 679
Db 66698 TGTAGTGTGATCATGCTCAGCTGCAACCTTGAACCCCTAGGCTCAAGTCACTCCTCCT 66639
QY 680 GCCTCAGCCTCCAAAGTACGTGGGACTACAACAGCTGTACCAATGCTGCTGCTGA----- 734
Db 66638 GCTTCAGCCTCCAAAGTACGTGGGACTATAGCATGTACCAATGCTGCTGCTGAATTTT 66579
QY 735 -----TATTTTCTTGTGAACAGGCTATCAGCTCTGTTGCCAGGCTGGAGTACAGTG 788
Db 66578 TAATTTCTTTTGTGAGCTGAGTCTGCTCTGTGCGCCAGGCTGGAGTCAAGTG 66519
QY 789 CGCTAATATAGCTACCTGCAGCCTCC--CCTCCTGGGCTCAAGCAATCCGCTGGCCTCAG 847
Db 66518 CGCGCATCTCGCTCAGCTGCAAGCTCCGCTCCCGGGTTCAGCCATCTCTCGCTCAA 66459
QY 848 CATCTAGTACGTGGGACTACAGCTGTGGCCACGACCCAGCCAGTAAAGTTTAAAAAT 907
Db 66458 CTTCCGAGTACGTGGGACTACAGTGTGGCCACGACCCAGCCAGTAAAGTTTAAAAAT 66399
QY 908 GATTTTGGTATAGAGAGGCTTCTGCTATGTTGCTCAGGCTGATTTTATTTGTTGAGAC 967
Db 66398 TATTTTGTAGACAGCGGGTTTCACTGTGTTAGCCAGGATGGTCTCGATCTCTGACG- 66340
QY 968 AAGTCTCAGTATGTGCGATGATCCCGCCACCTCCATCTCCCAAGTGTCTATCTATC 1027
Db 66339 -----TTGTGATCCACCTGCTCGGGCTCCCAAGTGTCTGAGATTACA 66297
QY 1028 TGTTCATTAGTACGTGACAGACATTTAGTGTGTTTCCACTTTTGGACATTAAGATAA 1087
Db 66296 GCGGTGAGCCACTGCACCCAGCCTAAATTTTAAATTTTCTGTAGACAGAGGTTTCGCTA 66237
QY 1088 TACTCCAGTGAATATCAATGATACATTTGTGCGGCATATGTTTCATTTCTGTTGGGT 1147
Db 66236 TGTGCCAGGCTGCTACTGAA-----CTACTGGCCTCAAGCATCTCTCTGCTTGG 66184
QY 1148 TTATATCTAGAGTGAATGCTGTAATCCCGGGTAATATTTTTCAGGACAGATTCAGGG 1207
Db 66183 CCACCAAAAGTCTGAGATTACAGSCATGAGTACCATCCCTCGGCTG----- 66136
QY 1208 GAAGAAAACTTGGGAAATGAAGCATGTTAGAAATCAGCAAGTGCAGGGGTTTTTC 1267
Db 66135 -----AATTTTGAATAAATGACAGTTTAAATTTTAAAGTTTATATGTAATTTATAC 66083
QY 1268 GGAGTTTATTTATATTCCTGTTGACAATGTCAGTTTGTATGATGATACAGTTATACT 1327
Db 66082 AGCTTAGTCTCTGAAATCAGCACTGAAGTGTAAATTTGTTATAGGTTGGAATCAAGT 66023
QY 1328 AAGTGAAGTGAATTAAGGCTGGAATGAGGCTTCCAGAGTAAATCATGAAGCACTT 1387
Db 66022 AAGAAATGTTAGAGACCACTACTATATCTTATATAGCAT-----TTT 65975
QY 1388 TGAATACAAATTAAGAGCTTGGCTGTAAACAAAATAAATAAATAAATAAATAAATAA 1447
Db 65974 TCTGTAATATAAACAACATTTTCCATATACAGGTTAAACCTCTGGGTTTTTTTGTGT 65915

```

```

QY 1448 TTTTCTTTTGTAGAAAGAGCTTTGCTTTTACCTGGCTGGAGGAGTGGTGTATCTC 1507
Db 65914 TTGTTTTTTTGTAGATGGAGTCTTGTGTCACCCAGGCTGGAGTGAATGGCAATCTC 65855
QY 1508 AGCTCAGTGAACCTTTGCTTCCCGGGTTCAAGCAATCTCTCTGCTGAGCTCCCAAGT 1567
Db 65854 GGCTCAGTGAACCTTCTGCTTCCCGGGTTCAAGTGAATCTCTGCTGAGCTCCCTGGAT 65795
QY 1568 AGCTGGGACTACAGGCACTTCCACCATGCCCAGCTGATTTTT--GTATTTTGTAGTAGA 1626
Db 65794 AGCTAGGACTACAGGCACTGCCCACCGCTGGCTCATTTTTCGTATTTTGTAGTAGA 65735
QY 1627 TGGGATTTCACTTTGTTGGCCAAAGCTGGTCTCAAACTTTTGTCTGTGATAAATTTGTA 1686
Db 65734 CGGGTTTTTACCATTTTACCCAGGCTGGTCTCAAACT--TCTTGACCTCAAGGATCTGCCT 65676
QY 1687 CTATTTGTTCTTTTCTGAGGTAGGCCCCCAGACAAAATAAATCTTGAATCCA 1746
Db 65675 GCCTCCTGCTCAGCTTCCAAAGTGTGGGATTTACGCCGTGAGCAACCATGCCAGCA 65616
QY 1747 AATCAGTGTGTTGTTGACCACTGTCTGAGAACCAACAGTGTGACCCAGGCTCAGG 1806
Db 65615 AAACCTTTTATTTTAT-----AATCAATTCAGATATCTCTAT----- 65579
QY 1807 AGTAGAGGTGATCTCTGCTCGAAAGAGAAATAGAAATGAAAATTTCTCCGGGCCAGGCT 1866
Db 65578 -GTGATTTCTCACTTTGTTTCAAAAAATTTGTGAAGAACTAAAGTATGCGCGGCGCAT 65520
QY 1867 GGTGCTCATGCTCTTAATCCAGCACTTTGGAGGCCAAGCATGTGGATCACTCAGG 1926
Db 65519 GGTGACTCATGCTGTATCCAGCACITTTGGAAGCCAGGTTGGTGGATCATTGAGG 65460
QY 1927 TCAGAGTTCAAACAGCCTGGCCAAACATGTTGTAACCCCTCTCTACTAAAAATACAA 1986
Db 65459 TTTGAGTTTGAGACCAACTGGCCAAACATGTTGTAACCCCTCTCTACTAAAAATACAA 65400
QY 1987 A-AAATTTAGTAAAGTGTGGTGGCATGCTGTATCCAGCTACTTGGAGGGTGGAGC 2045
Db 65399 AAAAATTTAGCCGGTGTGGTGGAGCACACCTGTAGTCTCAGTCTCAGGAGGCTGAGGC 65340
QY 2046 AGGAGATTTCTTTGAACCGGAGGAGGAGTGTGAGTGAAGCGAGATCACCACTGCA 2105
Db 65339 AGGAGATTTCTTTGAACCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 65280
QY 2106 CTCAGCCTGGGGAGAGAGGAGACTTCTCTCTCAAAAAACAAACAAACAAA 2158
Db 65279 CTCAGCCTGGGGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 65227

RESULT 12
AC004760/c 44496 bp DNA linear PRI 29-MAY-1998
LOCUS Homo sapiens chromosome 16, cosmid clone 316H7 (LANL), complete
DEFINITION sequence.
ACCESSION AC004760
VERSION AC004760.1 GI:3168626
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 44496)
AUTHORS Rieke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 44496)
Rieke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Misra,M. and Deaven,L.

```

[illegible]

repeat_region	complement(29311..29613)	QY	949	GTATTTTATTTGTTG-----AGACAAGGCTTCACATATGTTGCCATGATCCCCACCTC	1002
repeat_region	/rpt_family="Alu"	Db	30841	GGCAITCAACTCTTGGCTCAAGCAATCTCTGCTGCTTGGCTCTCTGAGTGTGCAAT	30782
repeat_region	30032..30332				
repeat_region	/rpt_family="Alu"	QY	1003	CACCTCCCAAGTGCATCTTATCTGTTCAATAGTACAGTACAGACATTTAGTGTGT	1062
repeat_region	30374..30700				
repeat_region	/rpt_family="L1MB7"	Db	30781	ACAGGCATGACCAACCGCAGCCCTGCTTCCACCTCTTGATGACCTTCAGATTGT	30722
repeat_region	30719..31200				
repeat_region	/rpt_family="Alu"	QY	1063	TCCACCTTTTGACCATATGAATAATACCTCCAGTGAATTCATGATATACATATGTTGG	1122
repeat_region	31038..31059				
repeat_region	/note="(A)22"	Db	30721	TTCAATTTTCGGGTATATTAATGAATGCTGCTAAGCAATTCATGTTAGGTGTTCTGT	30662
repeat_region	/rpt_type=tandem				
repeat_region	/rpt_unit=A	QY	1123	GCATATGTTTTCATTTCTGTTGGTTTATATCTAGAGTGGAAATGCTGATCCCGGTA	1182
repeat_region	31274..31826				
repeat_region	/rpt_family="Alu"	Db	30661	GGACATGTTTTCATGCTACAGGTATGCTCTCAGGTAGAACTGGGGTCTATATGTT	30602
repeat_region	31823..32146				
repeat_region	/rpt_family="L1MB7"	QY	1183	ATAT---TTTGACAGCAGATGTTGAGGGAAGAAACTTGGAAATGAGCATGTTTA	1239
repeat_region	32148..32444				
repeat_region	/rpt_family="Alu"	Db	30601	AATTCGGTTTGACTGTGTGAGGAGCTGCAAGATGTTTCTCGAGAGTTCGACCGTTT	30542
repeat_region	32441..32554				
repeat_region	/rpt_family="L1MB7"	QY	1240	GAATCAGCAAGAGTGCAGGG-----TTTTTCGGAGTTTATTTTATATTTCT	1287
repeat_region	complement(32676..32796)	Db	30541	TACATCTCACCAGAGTGTGGAGGGTCTGATTTTCCACATCTTAGTTATCCATTTT	30482
repeat_region	/rpt_family="Alu"	QY	1288	GT-TGACAAATGTGCAAGTTTGTATGAAGATACAACTATATCTAAGTGAAGTGAATTA	1346
repeat_region	complement(32836..32974)	Db	30481	GTATCACCTGTTTGTGTTGTTTCTTTAATTTTAAATATAGCCACCTTAGTGTGT	30422
repeat_region	complement(33162..33792)	QY	1347	AGGCTGGAATAGGGCTTCAGAGTAAATCATGAAGACATTTGAAATACCAAAATTAAGA	1406
repeat_region	34449..34471				
repeat_region	/note="(T)23"	Db	30421	AGGAAGTGTAACTCACTG-TAGTTTGTCTTGTGCTTCTTAACTCTCCAGGATCTTGA	30363
repeat_region	/rpt_type=tandem	QY	1407	GCTTGGCTGTAAACAAATAATAAAATACAAATTTTTTTTTTTTTTTTTTTTGAAGAA	1466
repeat_region	/rpt_unit=T	Db	30362	GTTTTTAAATAAATGATTTAGAGGAATATCTTTTTTTTTTTTTTTTTTTTGGAGGAGT	30303
repeat_region	complement(34450..34753)				
Query Match	10.0%; Score 415.2; DB 9; Length 44496;	QY	1467	CTGTCTTTTCCACCTGGCTGGAGGCGAGTGTGTGATCTCAGCTCAGCTGCAACTTTCG	1526
Best Local Similarity	58.8%; Pred. No. 8.4e-91;	Db	30302	CTGCTCTCTGCGCCAGGCTGGAGTGCAGTGGCGGGATCTCGCTCAGTGCAGCTCCG	30243
Matches 1044; Conservative 0; Mismatches 633; Indels 99; Gaps 15;		QY	1527	CTCCGGGTTCAAGCAATTTCTCTGCTTCCAGCTCCAGTCCAGTAGCTGGGACTCAGGCA	1586
		Db	30242	CTCCGGGTTCAAGCAATTTCTCTGCTTCCAGCTCCAGTCCAGTAGCTGGGACTCAGG	30183
		QY	1587	TCCACCATGCCCAGCTGAT-TTTTGTATTTTAGTAGAGATGGATTTTACATTTTGTGG	1645
		Db	30182	CACCAACACCCGGCTAATGTTTTTTTGTATTTTAGTAGAGCGGGTTTCCATGTTGG	30123
		QY	1646	CCAAGTGTCTCAAACTTTTTTGTCTCATATTTGTTAACTATTGTTCTCTTTTCTGTA	1705
		Db	30122	CCAGGATGCTCAAACTCT-----GAACTCGTGTATCCACTGC	30083
		QY	1706	GGTAGGCCCCCAGACCAAAATAATATCTTAGAATCCAAATCAGTGTGTTGTTGA	1765
		Db	30082	CTCAGCTCCCAAGTGTAGGATTAACAAGCTGAGCCACCGCGCCCGCTTAAGAGGA	30023
		QY	1766	CCACTGTCACTTTGAGAACACAGTGTGACCGGCTCAGGAGTGAAGTGTCTCTGCT	1825
		Db	30022	AATGTTTCTTAACAAATATATAACCAAGGCCATAAATATAGGACCAATAAAAT	29963
		QY	1826	CGAAAGAGAAATAGATGAAATATTTCTCCGGGCCAGGCTGTGGTCTCATGCTTAAT	1885
		Db	29962	TTAAGTCAAAAGACAACA-----CCCGGTGAGGTGCTCATGCTTAAT	29916
		QY	1886	CCAGCAGCTTTGGGAGGCCAAGCATGTGGATCCTCAGGTGAGGTGAGGTTCAAAACG	1945
		Db	29915	CCAGCAGCTTTGGGAGGCCAAGCATGTGGATCCTCAGGTGAGGTGAGGTTCGAGACG	29856
		QY	1946	CTGGCCCAACATGTTGAAACCCGCTCTACTAAAAATACAAAAATTTAGTGAAGTGTG	2005
		Db	29855	CTGGCCCAACATGTTGAAACCCGCTCTACTAAAAATACAAAAATTTAGTGAAGTGTG	29797

QY 2066 GCGCATCCCTGTAATCCAGCTACTTGGGAGGCTGAGCGAGGAGAAATTTCTTGAACCCG 2065  
 DB 29796 AGCGGGATCTGTAATCCAGCTACTTCCAGAGGCTGAGCGAGGAGACTTGTGCACTCG 29737  
 QY 2066 GGAGCGAGCTGTCAGTGAAGCGAGATCACACCACTGCATCCAGCTCCAGCTGGGGAGAGAG 2125  
 DB 29736 GGAGCGAGGCTGTCAGTGAAGCGAGATCGTGCCTACTGCATCCAGCTCCAGCTGGGCAACAGAC 29677  
 QY 2126 CGAGACTTCCTCTCAAAAAACAAAAAAGAA 2161  
 DB 29676 AGAGCGAGCTTACTCAAAAAAAGAAAAA 29641  
 RESULT 13  
 AC006111  
 LOCUS Homo sapiens chromosome 16 clone RP11-461A8, linear PRI 02-DEC-2000  
 DEFINITION Homo sapiens chromosome 16 clone RP11-461A8, complete sequence.  
 AC006111  
 VERSION AC006111.3 GI:11527451  
 KEYWORDS HYG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 190821)  
 AUTHORS Ricke,D.O.  
 TITLE Large Scale Sequence Analysis and Annotation with the Sequence  
 Comparison Analysis (SCAN) System  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 190821)  
 AUTHORS Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,  
 Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,  
 Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,  
 White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,  
 and Deaven,L.  
 TITLE Sequencing of Human Chromosome 16p13.3  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 190821)  
 AUTHORS Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,  
 Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,  
 Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,  
 White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,  
 and Deaven,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-DEC-1998) Center for Human Genome Studies, DOE Joint  
 Genome Institute, Los Alamos National Laboratory, MS M888, Los  
 Alamos, NM 87545, USA  
 REFERENCE 4 (bases 1 to 190821)  
 AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
 Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
 Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S.P.S.,  
 Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bussod,M.,  
 Sutherland,R. and Deaven,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-FEB-1999) Center for Human Genome Studies, DOE Joint  
 Genome Institute, Los Alamos National Laboratory, MS M888, Los  
 Alamos, NM 87545, USA  
 REFERENCE 5 (bases 1 to 190821)  
 AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
 Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
 Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S.P.S.,  
 Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bussod,M.,  
 Sutherland,R. and Deaven,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JUL-1999) Center for Human Genome Studies, DOE Joint  
 Genome Institute, Los Alamos National Laboratory, MS M888, Los  
 Alamos, NM 87545, USA  
 REFERENCE 6 (bases 1 to 190821)  
 AUTHORS Mundt,M.O.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-DEC-2000)  
 COMMENT On Dec 2, 2000 this sequence version replaced gi:5348385.  
 FEATURES Location/Qualifiers

## source

1. 190821  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-461A8"

BASE COUNT 48342 a 48030 c 48379 g 46070 t

## ORIGIN

Query Match 10.0%; Score 415.2; DB 9; Length 190821;  
 Best Local Similarity 58.8%; Pred. No. 1.1e-90;  
 Matches 1044; Conservative 0; Mismatches 633; Indels 99; Gaps 15;  
 QY 428 TAGCTCACTGAACCTCAAACTCCTGGGCTCAAGTGATCTACTACCTCAGCTCCAGCTCCAGAG 487  
 DB 163899 TTGGCCAGGCTAGTCTCAAACTCCTGACCTCAAGTAATCCACCCAGCTTAGCTCCCGAA 163958  
 QY 488 TAGCTAGACTACAGGCACACACAGCCATACCTGGCTAAATTTTTTTTTTAATTTTCATT 547  
 DB 163959 GTGCTGAGATTGTAGACATGAGCCACTGCACCCAGGCCAGCTTC-----CAAC 164006  
 QY 548 TTATGTAATTCATTTCTTTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 607  
 DB 164007 TTTTACGTAACTTTTGT 164066  
 QY 608 TACCAGGCTGGAGGCACTGCATGTGTGACAGCTGACAGGCTTGACTTCTCTGGGCTCA 667  
 DB 164067 AGTGGCTGTTTACAGGCACTGTCTATGTT-----GCATGTAGCTCCAACTGSCAGGCTTA 164122  
 QY 668 AGTGAATCCTCTGCTCAGCTCCCAAGTAGCTGGGACTACAACACAGCTGTCAACCATGCC 727  
 DB 164123 AGCNACTCTCCCACTCAGCTCCCAAGTAGCTGGGACTACAACACAGCTGTCAACCATGCC 164182  
 QY 728 TGGCTGATA-----TTTTTTTCTTTTAAACAGGGTATCACTCTGTGTG 769  
 DB 164183 TGGCTAATTTGCACTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 164241  
 QY 770 CCCAGGCTGGAGTACAGTGGGCTAATATAGCTCACTCACTCAGGCTC-CCCTCCTGGGCTCA 828  
 DB 164242 ACCAGGCTGGAATGAGTGGGCAACATCAGCTCAATAGCTCAAACTCCTGGGCTCG 164301  
 QY 829 AGCAATCCGCTGGGCTCAGCATCTGAGTAGCTGGGACTACAGGCTGTGCACACAGGCC 888  
 DB 164302 AGCAATCTCCCACTCAGCTCCCAATGCTGGGACCAAGGATGTGCCACACACC 164361  
 QY 889 CAGCTAATTTTAAATAAGATTTTGTGTATAGAGAGGCTTGTGTATGTGTGTGTGTGTGTGT 948  
 DB 164362 CAGCAATTTTAAAAA-----TTTTTTTTCGAGATGGAGTCTCACTATGTGTGTCCAGGCT 164416  
 QY 949 GTATTTTATTGTTG-----AGACAAGTCTCACTATGTGGCATGATCCCCACCTC 1002  
 DB 164417 GGCAATCAACTCTTGGGCTCAAGCAATCTCTGCTTGGGCTCTCTGAAGTGTCTCAAT 164476  
 QY 1003 CACTTCCCAAGTGTCTATCTTATCTGTCTAATAGTCTAGTTGACACAGATTTAGTGTGT 1062  
 DB 164477 ACAGGATGACCAACCGCACCCAGGCTGTTCACCTCTGTATGGACCTTCAGATTTGT 164536  
 QY 1063 TCCACTTTTTCACCATATGATATATATCTCCAGTGAATATTTCACTGTGTGTGTGTGTGT 1122  
 DB 164537 TTTCAATTTTCGGGTTTATGAAATATGCTGTATGAAATTTCACTGTGTGTGTGTGTGT 164596  
 QY 1123 GCATATGTTTTCAATCTGTTGGGTTTATATCTAGGAGTGAATTTGCTGGATCCCGGTA 1182  
 DB 164597 GGCAATGTTTTTCAATCTGTTACAGGTTATGCTCAGGTTAGACTGTGGGTCATATGTT 164656  
 QY 1183 ATAT---TTTCAGAGGAGAGTTTCAGGGGAAAGAAAACTTGGGAAATGAAGCATGTTTA 1239  
 DB 164657 AATTCGGTTTTCAGTGTGTGAGGAGCTGCAAGAATGTTTTCTCGCAGAGGTGCGACCGTTT 164716  
 QY 1240 GAAATCAGCAAGAGTGCAGGG-----TTTTTCGGAGTTTATTTTATATTTCT 1287  
 DB 164717 TACATTTCTCACCAGCAGTGTGGGAGGGTTCTGATTTTCTCCACATCTAGTTATCCATTTT 164776  
 QY 1288 GT-TCAGAAATGTGCAGTTTTCATGAAGATACAAGTTATATAGTGAGAGAGTGAATTA 1346

```

Db 164777 GTATCAGCTGTTTTTTTGGTTGTTCTTAATTTTAAATATATAGCCACCTAGTCAGT 164836
QY 1347 AGGCTGGAAATGAGCGGCTCAGAGTAAATCATGAAGCAGCTTTGTAATACCAAAATTAAGGA 1406
Db 164837 AGGAAGTGCTAACTCACTAGT-TAGTTTTGCTTTGAGTTCCTAACTCCAGGCACTCTGA 164895
QY 1407 GCTTGGCTGTAAACAATAATAAATAATACAAATTTTTTTTTTTTTTTTTTTTGGAAAAAGT 1466
Db 164896 GTTTTTAAAAAATATGTTTAAAGAGGAATCTTTTTTTTTTTTTTTTTTTTTTTTGGCGGAGT 164955
QY 1467 CTTCGCTTTCCACCTCGCTGGAGGAGGTGTGATCTCAGCTCAGCTCACTGCAACTTCGC 1526
Db 164956 CTCGCTGTGCGCCAGCTGGAGTGAGTGCGGGGATCTCGGCTCACTGCAAGCTCCGC 165015
QY 1527 CTCGCGGTTTCAAGCAATTTCTCTGCTTCCAGCTCCCAAGTAGTGGGACTTACAGGCACT 1586
Db 165016 CTCGCGGTTTCAAGCAATTTCTCTGCTTCCAGCTCCCAAGTAGTGGGACTTACAGGCACT 165075
QY 1587 TCCACCAATGCGGAGCTGAT-TTTTGTATTTTTAGTAGAGTAGGGATTTCACTTTGTGG 1645
Db 165076 CACCACACACCGCGGCTAAATTTTTGTATTTTTAGTAGAGCGGGGTTTCACTTGTGG 165135
QY 1646 CCRAGCTGGTCTCAAACTTTTGTGCTGTCATAATTTGTTCACTATTGTTCTTTTGTGTA 1705
Db 165136 CCAGGATGGTCTCAAACTTCTT-----GAACTCGTATCCACTTGC 165175
QY 1706 GGTAGGCGCCCGACCAAAAAAATAATCTTAGAATCCAAATCAGTGTGTTGTTTGA 1765
Db 165176 CTCAGCTCCCAAGTGTAGATTACAAGCTGAGCCACCGCGCGCTTAAGAGGA 165235
QY 1766 CCATGTCATCTGAGACCAACAGTGTGACAGGCTCAGGAGTAGAGGTGATCTCTGCT 1825
Db 165236 AATGTTCTTAAACAAATAATATAACCCAAAGGCCATAAATATAAGGACCAATATAAAAT 165295
QY 1826 CGAAGAGAAATAGATAAATAATTTCTCCGGCGCAGCGTGGGCTCATCGCTGTAAT 1885
Db 165296 TTAGTCAAAAGACAACA-----CCGGGTGAGGTGGCTCATCGCTGTAAT 165342
QY 1886 CCCAGCACTTTGGAGGCGCAAGGATGTGATCACTCAGTGCAGGAGTTCAAAACACAGC 1945
Db 165343 CCCAGCACTTTGGAGGCGTGAAGTGGTGGTATCGCTCAGTGCAGGAGTTCAAGACACAGC 165402
QY 1946 CTGCGCAATGTTGAACCCGCTCTCTACTTAAATAAATAAATAAATAGTAAAGTGTGCT 2005
Db 165403 CTGCGCAATGTTGAACCCGCTCTCTACTTAAATAAATAAATAAATAGTAAAGTGTGCT 165461
QY 2006 GCGCATGCTCTTATCCAGCTACTTTGGGAGGTGAGGAGGAGAAATTTCTTGAACCCG 2065
Db 165462 AGCGGGCATCTGTAATCCAGCTACTTCCAGAGGCTGAGGAGGAGAGCTTGTGCACTCG 165521
QY 2066 GGAGCGAGAGTTTCAGTGAAGCAGAGATCAACACCTCAGCTCAGCTGGGGGAGAGAG 2125
Db 165522 GGAAGCGAAGCTCAGTGAAGCAGAGATCGTCCACTCAGCTCAGCTGGGCAACAGAG 165581
QY 2126 CGAGACTTCTCTCAAAAAACAAAAACAAAAAGAA 2161
Db 165582 AGAGGAGACTCTACTCAAAAAAATAAAAAA 165617

```

## RESULT 14

```

AC091180/c  AC091180  183690 bp  DNA  linear  PRI 01-MAR-2002
LOCUS       Homo sapiens chromosome 17, clone RP11-1079K10, complete sequence.
DEFINITION  AC091180
ACCESSION  AC091180
VERSION    AC091180.20  GI:19033596
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 183690)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Homo sapiens chromosome 17, clone RP11-1079K10  
Unpublished  
2 (bases 1 to 183690)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Boguslavsky,L., Bouckghalter,B., Brown,A.,  
Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,  
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,  
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,  
Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRoque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., McPheters,R., Meldrim,J., Meneus,L.,  
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,  
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

Direct Submission  
Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS

3 (bases 1 to 183690)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckghalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,  
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,  
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,  
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL

Direct Submission  
Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE  
AUTHORS

4 (bases 1 to 183690)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckghalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,  
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,  
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,  
Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 1, 2002 this sequence version replaced gi:18874978.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L13058  
 Center clone name: 1079\_K\_10  
 -----  
 FEATURES  
 source  
 1. .183690 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="Rp11-1079K10"  
 /clone\_lib="RP11-11 Human Male BAC"  
 746..772  
 /rpt\_family="(A)n"  
 complement(2191..2292)  
 /rpt\_family="AluY"  
 2489..2522  
 /rpt\_family="AT-rich"  
 2749..3049  
 /rpt\_family="AluSg"  
 3560..3591  
 /rpt\_family="AT-rich"  
 3670..3834  
 /rpt\_family="FAM"  
 5596..5701  
 /rpt\_family="G-rich"  
 complement(6069..6199)  
 /rpt\_family="FLAM\_C"  
 complement(6459..6764)  
 /rpt\_family="AluSx"  
 complement(6769..7057)  
 /rpt\_family="AluJo"  
 complement(7277..7341)  
 /rpt\_family="MER208"  
 complement(7854..8057)  
 /rpt\_family="MIR"  
 complement(8387..8681)  
 /rpt\_family="AluSx"  
 8629..8965  
 /rpt\_family="FLAM\_C"  
 9054..9341  
 /rpt\_family="AluJb"  
 complement(9377..9671)  
 /rpt\_family="AluSc"  
 9764..10070  
 /rpt\_family="AluSx"  
 10128..10259  
 /rpt\_family="FLAM\_C"  
 complement(10332..10635)  
 /rpt\_family="AluSx"  
 complement(10981..11299)  
 /rpt\_family="AluY"  
 complement(11916..11994)  
 /rpt\_family="LTR66"  
 11954..12001  
 /rpt\_family="LTR66"  
 12027..12281  
 /rpt\_family="L2"  
 12446..12569  
 /rpt\_family="AluSg/x"  
 12684..12742  
 /rpt\_family="MIR"  
 12771..12874  
 /rpt\_family="L2"  
 complement(13139..13249)  
 /rpt\_family="MIR"  
 13326..13557  
 /rpt\_family="LIME"  
 complement(13558..13684)  
 /rpt\_family="FLAM\_C"  
 13685..13861  
 /rpt\_family="LIME"  
 complement(13862..14150)  
 /rpt\_family="AluSx"  
 14151..14226  
 /rpt\_family="LIME"  
 14246..14557  
 /rpt\_family="AluSx"  
 14558..14877  
 /rpt\_family="AluSg"  
 complement(15275..15573)  
 /rpt\_family="AluJb"  
 15619..15915  
 /rpt\_family="AluSx"  
 16309..16637  
 /rpt\_family="AluSx"  
 complement(17057..17350)  
 /rpt\_family="AluY"  
 17718..18021  
 /rpt\_family="AluJb"  
 complement(19258..19564)  
 /rpt\_family="AluSx"  
 complement(20258..20555)  
 /rpt\_family="AluSg"  
 20790..21102  
 /rpt\_family="AluSg"  
 complement(21034..21047)  
 /note"<30 qual SNGI region"  
 complement(21055..21059)  
 /note"<30 qual SNGI region"  
 complement(21064..21072)  
 Query Match 9.9%; Score 408.8; DB 9; Length 183690;  
 Best Local Similarity 57.5%; Pred. NO. 4.1e-89;  
 Matches 1281; Conservative 0; Mismatches 687; Indels 260; Gaps 20;  
 QY 4 AATTTCATCCATTTAAATCATACATACATTTAATGGCTTTTAGTATATATTCAC-AGGTTGTGCAT 62  
 DB 42197 AATTATCATTTTAAAGCATACAGTCTGCAGTTTGGTATATATTCACAGATTGTCCAA 42138  
 QY 63 CCATCAC-----AATCCATTTTGAACAGTTTATTTACTCCAAAATAAACCCGTCATCC 118  
 DB 42137 CCATCACCTATGTCTAATTCAGAACATTTTCATTCGCCCAAAAGAAAGCTGGAGCCCA 42078  
 QY 119 TTAGCATACCCCCCAACATCTCCATCTCTCCAGCCCTGGCAACACCAATCT 178  
 DB 42077 TCAGCAGTCACTCCCC-----CCTCCTCCCTACTCCAGCTCTGTTAATGA-----CG 42029  
 QY 179 ACTTTCGTCTCTATATAATTTGCCAATTCGTGGACATTCATATAATAATGAAGCAACAAC 238  
 DB 42028 ATTATCTTCTCTCTATTTGCTTATTCGCTGTTTCTATATAATAGATCATACAAC 41969  
 QY 239 ATGTGAGACT-TTGTGACTGSGCTTTCATCTAGTAGATCTTATTTTAAAGCTCAT-TAT 296  
 DB 41968 TAGTAGCCCTGTTATGCTGCTTCCTTTCACCTTAGCATAATGCTTTCGGGGTTTTCAT 41909  
 QY 297 GTTACAGTACTTAGCAGTACTTCTCTTTTATTTTATCTCAATGGTATTCCTCCTGCTGG 356  
 DB 41908 CTTGTACCATATATCATGACTACTACATTCCTTTTTT----- 41873  
 QY 357 GTATCCCATATCATATATTATAGACAGAGGTTCTCACCTCTGTCCACCGGCTGGAGTGCAG 416

Db 41872 -----TTTTTTGAGATCAGTCTTGTCTCTCCAAAGCTGGAGTGACG 41826  
QY 417 TGGCACAATCATAGTCTACTGTAACTCAAACTCCTGGCTCAAGTATCTACTACTC 476  
Db 41825 TGGCAAGATCTCAGTCTACTGCAACTCTGCTCTCTAGTTCGAAGCAATCTCCGCTTC 41766  
QY 477 AGCCTCCAGAGTAGTAGGATACAGGCACACACAGCATACCTGGCTAAATTTTTTTTT 536  
Db 41765 AGCCTCTCAAGTAGTGAACACAGGTGCATGCCACCACTCTCAGCTAAATTTGTATT 41706  
QY 537 TAAATTTTCAATTAATGATCAATTTCTTTCTTTTCTTTTGTGTTGTTTGGAGATAGG 596  
Db 41705 -----TTTGGTAGAGTGGG 41690  
QY 597 TCTCACTTTGTTTACCCAGCTGGAGGGCAGTGGCATGTGTGACAGCTGAGCAGCCTTGACT 656  
Db 41689 TTTCAACATGTTGCCACGCTG-----GTCTCGAAC 41659  
QY 657 TCCGTGGGCTCAAGTGATCCTCCTGCTCAGCCTCCCAAGTAGCTGGGACTACAACACGT 716  
Db 41658 TCCGTGAACCTAGCAATTCACCTGCTTAGCCTCCGCTAGTTCTGGGATTTACAGGCCATA 41599  
QY 717 GTCACCATGCCCTGCTGATATTTTTTTTCTTGAACAGGGTATCACTCTGTGGCCAGGC 776  
Db 41598 GTCACCATGCCCGGC--CATTTCTTTTGTGAGACAGGGTCTCACTCTGTGGCCAGAC 41541  
QY 777 TGGAGTACAGTGGCGTAAATATAGTCACTGACAGCTC--CCCTCCTGGGCTCAAGCAATC 835  
Db 41540 TGGAGTGTAGTGAATGATCATGGCTCACTGACAGCTCAACTCCAGGCTCAAGCAATA 41481  
QY 836 CGCTGGCTCAGCATCTCTGAGTGTAGTGGGACTTACAGGCTTGTGGCCAGGCCAGCTAA 895  
Db 41480 TTCCACACCTTAGCCCTCTGAGTGGCTAGGACTACAAATGTGCACCGCTACACCCAGCTAA 41421  
QY 896 GTTTTAAAAAATGTTTTTGTATAGAGAGGTCTGTGTATGTTGCTCAGGCTGTATTTT 955  
Db 41420 TTTTGTAAAA--ACTTTTGTAGAGAGGAGGTCTGTGTATGTTGACCAAGGCTGTGCA 41364  
QY 956 TATTTGTTGAGACAAGGTCTCACTATGTGGCATGATC----- 992  
Db 41363 AACTCTGGGCTCAGTATCATCCACACAGCCTCCCAAGTCTGGGATACAGCA 41304  
QY 993 -----CCCCCAGCTCCAGTCCCAAGTCTCACTATCTGTTCATTTAGTCAAGTGTACAG 1048  
Db 41303 TGAGCCACTACACCTGACTTCTCTCATCTCTTTTATAGACATAATATCCATTTGTATGA 41244  
QY 1049 ACATTTAGTTGTTT-----CCACTTTTGGACATTT 1079  
Db 41243 TTATATATTTTGTGTTTATCCACTCATCATGTTGAIGTATATTTGGATGTCCTTTTGGCTATT 41184  
QY 1080 ATGAATAATCTCCAGTGAATATTCATGTATACATTTGTGTGGGCATATGTTTTCATTTTC 1139  
Db 41183 ATAAATATGCTCTGTGAACATTTGTTTACAAGTTTTCGATGAACATGTTTGTGTTTC 41124  
QY 1140 TGTGCGTTTATATCTAGAGTGGAAATGCTGGATCCCGGGTAAATATTTTGACAGGCAGA 1199  
Db 41123 TCTCAGATTTATGCTGGAACGTGAATTTCTGAGTCAATATGTTAACTGTTTAAATTTTTTT 41064  
QY 1200 GTTCAGGGGAAGAAAACCTTGGGAAATCAAGCATGTTTAGAAATCAGCAAGAGTGCAGG 1259  
Db 41063 TTTTAAAGNACTGCCAAACTGTTCTCCAGTGTACCATTTTATATCTCCAGGTAACATA 41004  
QY 1260 GGTGTTTCGGAGTTTATTTTATATTC-----TGTGACAAATGTGCAGTTTGTATGAAGAT 1315  
Db 41003 TGAGGCTTCCATATCTTCACATCCTCCACACACTGTGCTGTGTTCCAAATTTTATAT 40944  
QY 1316 ACAAGTTATCTAAGTGAAGTGAAGATTAAGGCTGGAAATAGGCGTTTACAGAGTAAAT 1375  
Db 40943 AGCCATCTCTAGTAGGTGTGAAGTGGTATCTCATTTGAGATTTTG---ATTGCAATTTCACT 40887  
QY 1376 CATGAGACCTTTGATACCAAAATTAAGGAGCTTGGCTGTAAACAAATAATAATAAAT 1435

Db 40886 RATGCTGAATATGTTAAGCATCTTTTCTTTACGCAITGGCAITTTGTATATCTCTTTT 40827  
QY 1436 CACAATTTTTTTTTTTTTTTTGGAGAAAGAGTCTGTCTCTTTTACCTGTGCTGGAGGACG 1495  
Db 40826 TTTTTTTTTTTTTTTTTTTCTTGAGACAGAGTCTCACTGTGCGCCAGGCTGAGTGCAG 40767  
QY 1496 TGGTGTGATCTCAGCTCACTGCAACTTTTGCCTCCCGGTTTCAAGCAATTTCTCTGCTTC 1555  
Db 40766 TGGCAGGATCTCGGCTTACTGCAAGCTCGGCTCTCTGGTTCACACCAATTTCTTCTGCTC 40707  
QY 1556 AGCCTCCCAAGTAGTGGGACTACAGGCACTTCCACACANGCCAGCTGATTTTTTG--- 1611  
Db 40706 AGCCTCCCAAGTAGTGGGACTACAGGCACTTCCACACANGCCAGCTGATTTTTTTT 40647  
QY 1612 -----TATTTTAGTAGAGTGGGATTTTCACTTTTGTGGCAAGCTGGTCTCAAACT 1663  
Db 40646 TTTTTTTTTTTTTTTTTAGTAGACAGGGTTTCACTATGTTAGCCAGGATGGTCTCGATCT 40587  
QY 1664 TTTTCTGTCAATAATTTGTGTAACTATTGTTCTTTTGTGAGGTAGGCCCCCAGACCA 1723  
Db 40586 CCTGAC-----CTTGTGATCCACCCACTTGGCTCCCAAAGTCTGGGATTACA 40537  
QY 1724 AAAAAAATAAATCTTAGAATCCAAATCAGTGTGTTGGTTGACCACTGTCACTTGAAGAC 1783  
Db 40536 GGTGTGAGCCACACGCGCTGGCCCATTTGTATATCTCTTTAGAAATGTTCTTCAATTT 40477  
QY 1784 CACAGTGTGACCCAGGCGCTCAGGAGTAGAGTGATCTCTGCTCGAAAGAGAAATAGAATG 1843  
Db 40476 CTTTCCGT-----TTTAAATTTTTTAA 40456  
QY 1844 AAAATATTCTCCGGGCCAGGCGTGTGGCTCATGCGTGPAAATCCAGCACTTTGGGAGCG 1903  
Db 40455 ATTATTTATTTTGGCGGGTGCAGTGGCTCATGCTGTAATCCAGCACTTTGGGAGCG 40396  
QY 1904 CAGGCACTGTGATCACTCAGGTGAGGTTCAGAGTTCAAAACAGCCTGCGCAACATGTTGAAA 1963  
Db 40395 CGAGTGGGCGGATCAC--GAGGTTGGGAGATCGAGACCATCTCTGCTAACAGGTGAAA 40338  
QY 1964 CCCCCCTCTCTACTAAAAATACAAAAATTAGCTAAGTGTGTGGCGCATGCTCTGTAATCC 2023  
Db 40337 CCCCCCTCTCTACTAAAAATACAAAAATTTAGCTGGCATGGTGACAGGCACTGTAGTCC 40278  
QY 2024 CAGCTACTTGGAGGCTGAGGAGAGAGAAATTTCTTGAACCCGGGAGGAGAGGTTGCACT 2083  
Db 40277 CAGCTACTCGGAGGCTGAGGAGAGAGAAATGCGTGAACCCAGGAGCGGAGCTTGCAT 40218  
QY 2084 GAAGGAGATACACACTGCACTCCAGCTCGGGGAGAGAGAGAGCTCTCTCTCAAAA 2143  
Db 40217 GAGCTGAGATCGCGCACTGCACTCCAGCTGGGCGACAGAGGAGACICTGTCTCAAAA 40158  
QY 2144 AAACAAAA 2151  
Db 40157 AAAAAAAA 40150

## RESULT 15

AC083866

LOCUS

DEFINITION

AC083866

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC083866 Homo sapiens chromosome 7 clone RP11-702D16, complete sequence.  
170528 bp DNA linear PRI 28-NOV-2000

AC083866.2 GI:11386323

HTG.

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 170528)

Kaul, R.K., Zhou, Y., James, R.A., Raymond, C., Haugen, E.D. and

Olson, M.V.

Large-scale Mapping and Sequencing of Human Chromosome 7

Unpublished

2 (bases 1 to 170528)

Kaul, R., Olson, M.V., Raymond, C., Clendenning, J. and Haugen, E.

BglII				EcoRI				HindIII			
SeqDerMap	FngRPrint	SeqDerMap	FngRPrint	SeqDerMap	FngRPrint	SeqDerMap	FngRPrint				
6121	6298	8696	8797	3998	3805						
2067	2067	6	<800	6382	6481						



```

FEATURES
source      Location/Qualifiers
-----
3775       3781       642       <800
-----
986       1023
-----
511       <800
-----
3242       3212
-----
2038       2046
-----
1. 170528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-702D16"
/sex="male"
/tissue_type="blood"
/clone_lib="rPCI human BAC library 11"
complement(17941..18218)
/standard_name="swSS3783"
43178..43179
159989..160063
/standard_name="swSS2778"
BASE COUNT 44972 a 35957 c 37542 g 52057 t
ORIGIN

Query Match      9.8%; Score 407.4; DB 9; Length 170528;
Best Local Similarity 58.6%; Pred. No. 8.8e-89;
Matches 1078; Conservative 0; Mismatches 631; Indels 132; Gaps 16;

QY 358 TATCCATATATATATAGAGCAGGTTCTCAGTCTGTCAACCAGGCTGGA-GTGCAG 416
Db 137212 TATTATTATTATTATTGAGACAGGCTTGTTCTGTCAACCAGGCTGGAGGTGCAG 137271

QY 417 TGGCAATATAGTCACTGTAACTCAAACTCCTGGGCTCAAGTGATCCTACTACCTC 476
Db 137272 TAGCATGATCATAGTCACTCAGCTTCAACCTCCTGGCTCAATTGATCTCCACCTC 137331

QY 477 AGCTCCAGAGTAGTAGGACTACAGGCACACAGCAGCATCCTGGCTAAATTTTTTTT 536
Db 137332 AGCTCCAGTAGTAGGACTACAGGCATGAGGATGACACACAGCAGCTGGCTAAATTTTTTTT 137390

QY 537 TAAATTTTCAATTTATGATATCTTCTTTTCTTTTCTGTGTGTGTTTTCAGATAGG 596
Db 137391 -----GTTTGTGTTGTTATTGTTAGATAGAG 137419

QY 597 TCTCACTTTGTTACCCAGGCTGGAGGAGTGGCATGGTGACAGCTGAGCAGCCTTGACT 656
Db 137420 TTTTACCGTGTGCCAGGCTG-----GTCCTGAAC 137450

QY 657 TCCTGGGCTCAAGTGATCCTCTGCTCAGCCTCCCAAGTAGCTGGGACTACAACACAGT 716
Db 137451 TCCTGGGCTCAAGCAGCATGCGCCAGCTCTGCTTCCAAAGTGTGGGATTTACAGGTAAA 137510

QY 717 GTCACCATGCTGGCTGAT-----ATTTTTTTCTTGAACAGGGTATCACTCT 765
Db 137511 GCCACCAACAGGGCTATTTTAATTTTATTTTGTAGTCAAGGGCTTGCTGT 137570

QY 766 GTTGCCAGGCTGGAGTACAGTGGGTAAATAATAGCTCACTGACGCCCTC-CCTCCTGGG 824
Db 137571 GTCACCCAGGCTGGAGTGCAGTGGGCACATATACACTCACTGACGCCCTCACTCCTGGG 137630

QY 825 CTCAGCAATCCGTCGGCTCAGCATCCCTAGTAGTCTGGGACTACAGGCTGTGTCACCA 884
Db 137631 CTCAAAGTATCTCTGCTCAGCCTGCTGCTAATAGTACAGGACAGGCTGTGTCACAA 137690

QY 885 GCCCAGCTAAGTTTTAAAAATGATTTTGTATAGAGGAGTCTTGTCTATGTTGCTCA 944
Db 137691 TGCCCTAGCTAATTTTAAAT-----TTTTTTGTAGAGTGGGATCTCACTATGTTGCTTA 137746

QY 945 GGCCTGATTTTATTGTTGAGACAAAGGCTCACTATGTTGCCATGATCCCCCACTCCA 1004

Db 137747 GGCTGCTTGAACCTCTGGGT-----TCAGCAGTCCCTTGCCCTGGCTCTCCA 137799
QY 1005 CTTCCTCAAGTCTCACTTAATCTGTTTCATTAGTCAAGTGCAGACATTTAGTGTTC 1064
Db 137800 GTGTTGAGATTGCGGGCGCAAGCCACCTCGTCCAACCTGTCTCTTTTGTGTGATTTAG 137859

QY 1065 CACTTTTTGACCATATATGAATAAATACCTCCAGTGAATATCATGTATACATTTGTGTGGC 1124
Db 137860 TAACATTTGATGATTTTACTTAGATCCATTAATCATTAAGAGATATACAAATGATGATA 137919

QY 1125 ATATGTTTTCTATCTGTTGGTTTATATCTAGGAGTGAATTTGCTGGATCCCGGTAAAT 1184
Db 137920 TTCTAGTTCTATTATCTTTTCATTATTATTAGCTGGAACTCTTA-----TAAA 137967

QY 1185 ATTTTGACAGCAGAGTTTCAGGGGGAAGAAACATTTGGGAAATGAAGCATGTTTGAAT 1244
Db 137968 GAGAAACCTCCCTTTTCTAATCTTCGATTTACCCAGTGGTATAATACATATTTGAAAGT 138027

QY 1245 CAGCAAGAGTGCAGGGGTTTTTCGGAGTTTTTATTTATATCTGTTGACAAATGTCAGT 1304
Db 138028 CAGGATGAATATTTTGTCTTTTCCCTGTTTACCAATTTTCCAGAACTCTGTTATGGTTC 138087

QY 1305 TTGATGAAGATACAAAGTTATATACTAAGTGAGAAATGAAGCTGAATAGGCGCTT 1364
Db 138088 ACTAGCATTCCTCAA-TTAGGCATTTGTTTGGTGGGGTTTTTTTGTATCATAACTC 138146

QY 1365 CAGATAAAATCATGAAGCAGCTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAA 1424
Db 138147 TTGGGTTTAAACATTTTGAATATGTTTCAATTCATTCATGCTTTTTCCTTTCTAGTGA 138206

QY 1425 TAAATAAAA-----ATCACAAATTTTTTTTTTTTTTTTGGAGAGAGCTTGTCTCTTC 1477
Db 138207 TTTTACTGATATATTTCTTTCTTTCTTTTGTGACGCGAGCTTGTCTCTGT 138266

QY 1478 ACCCTGGCTGGAGGAGCTGTGTGATCTCAGCTCACTGCAACTTTGCGCTCCCGGTTTC 1537
Db 138267 GCCCAGGCTGGAGTGCAGTGGTGCAATCTCAGCTAACTGCAACCTCCGCTCTAGGGTTC 138326

QY 1538 AAGCAATTTCTCTGCTTCCAGCTCCCAAGTAGCTGGGACTACAGCAGCTTCCACCATGC 1597
Db 138327 AAGTGATTTCTCTGCTCAGCTCCTGAGTAGCTGGGATACAGCCATGACACCTCGC 138386

QY 1598 CCAGCTGATTTTTGTATTTTGTAGAGATGGGATTTCACTTTGTTGGCAAGAGTGGCT 1657
Db 138387 CCAACT-AAATTTGTATTTTGTAGTAGAGTGGGTTTCTCCATGTTGTGTCAGCTGGTCT 138445

QY 1658 CAAACTTTT-----TGCTGTCAATAATTTGTTAACTATTTGTTTCTTTTGTGAGTAGGGC 1713
Db 138446 CGAACTCTTGACCTCAGGTGATCTCTCCGCTCGGCTCCCAAGTACTAGGATTACAGG 138505

QY 1714 CCCAGACCAAAAAAATAAATCTTTAGAAATCCAAATCAGTGTGTTGTTGACCACTGTC 1773
Db 138506 CACCATGCCAGTGGCATATATATTTCTTAAATAACTCAGCTGTGATGATGAGGACGTG 138565

QY 1774 ACTTGAGAACCACTGTCACCCAGGCGCTCAGAGGTAGAGGTGATCTCTGCTGCAAGAG 1833
Db 138566 ICTTACATCTTTCTGCTCCATGTTGCTAGTGGCTGGCTTTCAATAATGTCAAATGAA 138625

QY 1834 AATAAGAAAT-----GAAAAATTTCTCCGGGCCAGGGGTGGTCTCA 1875
Db 138626 TGAGTTGATATTTCTCAGCAGCATTTTAAAAATAAGAGCTAGTGGCTGGGCACGCTGCTCA 138685

QY 1876 TGCTGTGTAATCCAGCAGCTTTGGGAGGCCAAGGCATGTTGATCAGCTGAGGTGAGGAT 1935
Db 138686 TGCTGTGTAATCCAGCAGCTTTGGGAGGCCAAGGCAGGCGGGGATGATCAC--GAGGTGAGGAT 138743

QY 1936 CAAAACAGCCTGGCCAACTGTTGAAACCCGCTCTCTACTAAAAATACAAAAAATAGC 1995
Db 138744 -GAGACCAATTCGGCCCAACATGTTGAAACCCCTGCTCTACTAAAAATAC-AAAAATTAGC 138801

QY 1996 TAAAGTGTGTGGCGCATGCTCTAATCCAGCTACTTGGAGGGGTGAGCAGGAAATTT 2055

```

Db 138802 CGGAGTGTGTGGTGCCTTTAGTCCAGCTACTCGGAGGCTGAGCGAGGAATGG 138861  
QY 2056 CTTGAACCCGGGAGGCGAGAGTTGTCAGTGAAGCGAGATCACACCACTGCACCTCCAGCCTG 2115  
Db 138862 TATGAACCCGGGAGGCGGAGCTTGCAGTGAGCTGAGATTGGCCACTGCACCTCCAGCCTG 138921  
QY 2116 GGGGAGAGCGGAGACTTCCTCTCAAAAAACAAAAACAA 2156  
Db 138922 GCGACAGAGTGAGACTCTGTCTCAAAAAACAAAAACAA 138962

Search completed: May 20, 2003, 19:44:35  
Job time : 18194 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 12:50:07 ; Search time 820 Seconds  
(without alignments)  
11397.309 Million cell updates/sec

Title: US-10-016-725-15

Perfect score: 4150

Sequence: 1 aggaattccatttaaat.....tgagtttccagctggccc 4150

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

	1:	2:	3:	4:	5:	6:	7:	8:	9:	10:	11:	12:	13:	14:	15:	16:	17:	18:	19:	20:	21:	22:	23:	24:
	/SID22/gcgdata/geneseq/geneqseq-emb1/NA1980.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1981.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1982.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1983.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1984.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1985.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1986.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1987.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1988.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1989.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1990.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1991.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1992.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1993.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1994.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1995.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1996.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1997.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1998.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA1999.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA2000.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA2001A.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA2001B.DAT.*	/SID22/gcgdata/geneq/geneqseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4150	100.0	4150	24	ABL57445
2	1932.8	46.6	1970	24	Human stearoyl-CoA
3	612.4	14.8	617	22	Human stearoyl-CoA
C 4	429	10.3	17704	23	Human stearoyl-CoA
C 5	427.4	10.3	17707	23	Genomic sequence #
C 6	386	9.3	8680	22	Genomic sequence #
C 7	375.4	9.0	92638	22	Human immune/haema
C 8	368.8	8.9	5491	22	Human osteoblast d
C 9	366.4	8.8	9742	22	Human immune/haema
					Genomic sequence #

10	356.2	8.6	84607	20	AAx90847	Human PACAP genomi
11	349.8	8.4	12394	22	AAAD14749	Human glycoen syn
12	349.2	8.4	377	21	AAAC04246	Human secreted pro
13	349	8.4	13216	22	AAAL05122	Human reproductive
14	349	8.4	13216	23	ABL98014	Human testicular a
C 15	348.6	8.4	396	23	ABV52505	Human prostate exp
16	347.4	8.4	29163	22	AAAL05121	Human reproductive
17	347.4	8.4	29163	23	ABL98013	Human testicular a
C 18	347	8.4	6437	22	AAAS20588	Genomic sequence #
C 19	345	8.3	84495	24	AAAS20588	Human methionine a
C 20	343.2	8.3	128978	24	ABK83459	Human cDNA differe
C 21	342.8	8.3	26928	20	AAZ32184	Human prothrombin
C 22	342.8	8.3	26928	24	ABN95780	Gene #2278 used to
C 23	337	8.1	109906	24	ABK94411	DNA encoding endot
24	335.2	8.1	128978	24	ABK83459	Human cDNA differe
25	332	8.0	25715	22	AAAS33462	DNA encoding human
C 26	329.6	7.9	9372	22	AAAL36684	Human musculoskele
C 27	329.6	7.9	9372	22	AAAL36684	Human musculoskele
C 28	328	7.9	8308	22	AAK68091	Human immune/haema
C 29	328	7.9	8308	22	AAK79820	Human immune/haema
C 30	327.6	7.9	17397	22	AAAS36445	Human cardiovascular
C 31	327.6	7.9	19334	22	AAAS36443	Human cardiovascular
C 32	327.6	7.9	19345	22	AAAS36444	Human cardiovascular
C 33	327.6	7.9	25423	22	AAK90279	Human digestive sy
C 34	327.6	7.9	25423	22	AAI57656	Human colorectal c
C 35	327	7.9	5932	22	ABA07891	Human ovarian and
C 36	327	7.9	5932	22	AAAL03708	Human reproductive
C 37	326.6	7.9	18010	22	AAAK67807	Human immune/haema
C 38	326.4	7.9	32148	22	AAAL04218	Human reproductive
C 39	326	7.9	25424	22	AAK90280	Human digestive sy
C 40	326	7.9	25424	22	AAI57657	Human colorectal c
C 41	325.4	7.8	32145	22	AAAL04631	Human reproductive
C 42	325.4	7.8	32145	23	ABL97544	Human testicular a
C 43	325.2	7.8	24908	22	AAK81665	Human immune/haema
C 44	325.2	7.8	24999	22	AAK65622	Human immune/haema
C 45	325.2	7.8	48727	22	AAK67375	Human immune/haema

#### ALIGNMENTS

##### RESULT 1

ABL57445

ID ABL57445 standard; DNA; 4150 BP.

XX ABL57445;

XX AC

XX DT

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

XX DE

```

FT /bound_moiety= "NF-Y"
FT complement (255..258)
FT /*tag= f
FT /bound_moiety= "NF-Y"
FT complement (263..266)
FT /*tag= g
FT /bound_moiety= "IRF-2"
FT complement (461..464)
FT /*tag= h
FT /bound_moiety= "SREBP"
FT complement (541..544)
FT /*tag= i
FT /bound_moiety= "IRF-1"
FT complement (669..672)
FT /*tag= j
FT /bound_moiety= "SREBP"
FT 991..994
FT /*tag= k
FT /bound_moiety= "SREBP"
FT complement (1037..1040)
FT /*tag= l
FT /bound_moiety= "AP-1"
FT complement (1073..1076)
FT /*tag= m
FT /bound_moiety= "RAR-alpha"
FT 1240..1243
FT /*tag= n
FT /bound_moiety= "C/EBP-beta"
FT 1410..1413
FT /*tag= o
FT /bound_moiety= "NF-1"
FT complement (1423..1426)
FT /*tag= p
FT /bound_moiety= "HNF3-beta"
FT complement (1500..1503)
FT /*tag= q
FT /bound_moiety= "SREBP"
FT complement (1587..1590)
FT /*tag= r
FT /bound_moiety= "NF-kB"
FT 1628..1631
FT /*tag= s
FT /bound_moiety= "NF-kB"
FT complement (1632..1634)
FT /*tag= t
FT /bound_moiety= "IRF-1"
FT 1713..1716
FT /*tag= u
FT /bound_moiety= "AP-2"
FT complement (1729..1732)
FT /*tag= v
FT /bound_moiety= "HNF3-beta"
FT 1833..1836
FT /*tag= w
FT /bound_moiety= "IRF-2"
FT complement (1863..1866)
FT /*tag= x
FT /bound_moiety= "Whn"
FT 1917..1920
FT /*tag= y
FT /bound_moiety= "SREBP"
FT 1925..1928
FT /*tag= z
FT /bound_moiety= "T3R, RAR-alpha"
FT complement (2087..2090)
FT /*tag= aa
FT /bound_moiety= "Pax-6"
FT 2093..2096
FT /*tag= ab
FT /bound_moiety= "SREBP"
FT complement (2116..2119)
FT /*tag= ac
FT /bound_moiety= "AP-2"
FT

FT protein_bind
FT complement (2179..2182)
FT /*tag= ad
FT /bound_moiety= "CHOP"
FT 2199..2202
FT /*tag= ae
FT /bound_moiety= "RAR-alpha"
FT complement (2231..2234)
FT /*tag= af
FT /bound_moiety= "NF-1"
FT 2245..2248
FT /*tag= ag
FT /bound_moiety= "Pax-6"
FT 2341..2344
FT /*tag= ah
FT /bound_moiety= "RAR-alpha"
FT 2359..2362
FT /*tag= ai
FT /bound_moiety= "NF-1"
FT complement (2365..2368)
FT /*tag= aj
FT /bound_moiety= "ER"
FT complement (2382..2385)
FT /*tag= ak
FT /bound_moiety= "NF-1"
FT complement (2414..2417)
FT /*tag= al
FT /bound_moiety= "Rel"
FT 2445..2350
FT /*tag= am
FT /bound_moiety= "AP-1"
FT complement (2452..2455)
FT /*tag= an
FT /bound_moiety= "NF-1"
FT complement (2533..2536)
FT /*tag= ao
FT /bound_moiety= "GR"
FT complement (2726..2729)
FT /*tag= ap
FT /bound_moiety= "CREB"
FT complement (2744..2747)
FT /*tag= aq
FT /bound_moiety= "ER"
FT 2862..2865
FT /*tag= ar
FT /bound_moiety= "NF-1"
FT complement (2917..2920)
FT /*tag= as
FT /bound_moiety= "PPAR-alpha"
FT complement (2942..2945)
FT /*tag= at
FT /bound_moiety= "Oct-1"
FT 2958..2961
FT /*tag= au
FT /bound_moiety= "CREB/ATF"
FT complement (2965..2968)
FT /*tag= av
FT /bound_moiety= "STAT"
FT complement (2979..2982)
FT /*tag= aw
FT /bound_moiety= "C/EBP-beta"
FT 3056..3059
FT /*tag= ax
FT /bound_moiety= "NF-kB"
FT complement (3062..3076)
FT /*tag= ay
FT /bound_moiety= "NF-kB"
FT complement (3065..3068)
FT /*tag= az
FT /bound_moiety= "SP-1"
FT 3131..3134
FT /*tag= ba
FT /bound_moiety= "Ets-1"
FT complement (3161..3164)
FT

```

```
FT      /*tag= bb
FT      /bound_moiety= "Whn"
FT      complement (3162...3166)
FT      *tag= bc
FT      /bound_moiety= "Myc/Max"
FT      3173...3176
FT      *tag= bd
FT      /bound_moiety= "sp-1"
FT      3235...3238
FT      *tag= be
FT      /bound_moiety= "NF-kB"
FT      complement (3241...3244)
FT      *tag= bf
FT      /bound_moiety= "NF-kB"
FT      complement (3245...3248)
FT      *tag= bg
FT      /bound_moiety= "Whn"
FT      complement (3276...3279)
FT      *tag= bh
FT      /bound_moiety= "NF1"
FT      3291...3294
FT      *tag= bi
FT      /bound_moiety= "Myc/Max"
FT      3292...3295

Query Match      100.0%; Score 4150; DB 24; Length 4150;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGGAATTCATCCATTAAATCATACAAATTAATGGCTTTAGTATATTCACAGGTTGGC 60
DB      1 AGGAATTCATCCATTAAATCATACAAATTAATGGCTTTAGTATATTCACAGGTTGGC 60

QY      61 ATCCATCAAAATCCATTTTAGACAGTTTATTACTCCAAAATAAACCCTGCATTCCTT 120
DB      61 ATCCATCAAAATCCATTTTAGACAGTTTATTACTCCAAAATAAACCCTGCATTCCTT 120

QY      121 AGCCATCACCCCAACATCCCTCCATCCCTCCAAAGCCCTGGGCAACCAACAACTCTAC 180
DB      121 AGCCATCACCCCAACATCCCTCCATCCCTCCAAAGCCCTGGGCAACCAACAACTCTAC 180

QY      181 TTTCGTCTCTATAATTTGCCAATCTCGGACATTTTCATATAAATGGAAGCAACAAAT 240
DB      181 TTTCGTCTCTATAATTTGCCAATCTCGGACATTTTCATATAAATGGAAGCAACAAAT 240

QY      241 GTGAGACTTTGTGACTGGCTGCTTTTACATAGCATTTCTATTTTAAAGGCTCATTTATGTTA 300
DB      241 GTGAGACTTTGTGACTGGCTGCTTTTACATAGCATTTCTATTTTAAAGGCTCATTTATGTTA 300

QY      301 CAGTACTTAGCAGTACTTCATCTTTTATTTCTCAAATGGGTATTCACCTGTGGGTAT 360
DB      301 CAGTACTTAGCAGTACTTCATCTTTTATTTCTCAAATGGGTATTCACCTGTGGGTAT 360

QY      361 CCCATATCATATTTAGAGACAGGTTCTCATCTGTACCCAGGCTGGAGTGCAGTGGC 420
DB      361 CCCATATCATATTTAGAGACAGGTTCTCATCTGTACCCAGGCTGGAGTGCAGTGGC 420

QY      421 ACAATCATAGCTCACTGTAACTCAAACCTCCGGGCTCAAGTGATCCTTACTACTCAGCC 480
DB      421 ACAATCATAGCTCACTGTAACTCAAACCTCCGGGCTCAAGTGATCCTTACTACTCAGCC 480

QY      481 TCCAGAGTAGTAGGACTACAGGCACACAGCCATACCTGGCTAAATTTTTTTTTTAAT 540
DB      481 TCCAGAGTAGTAGGACTACAGGCACACAGCCATACCTGGCTAAATTTTTTTTTTAAT 540

QY      541 TTTCAATTTATGATTTCAATTTCTTTTCTTTTGTGTTGTTTGTAGATAGGCTTC 600
DB      541 TTTCAATTTATGATTTCAATTTCTTTTCTTTTGTGTTGTTTGTAGATAGGCTTC 600

QY      601 ACTTTGTTACCCAGGCTGGAGGCACTGGCATGGTGACAGCTGAGCAGCCTTGACTTCCT 660
DB      601 ACTTTGTTACCCAGGCTGGAGGCACTGGCATGGTGACAGCTGAGCAGCCTTGACTTCCT 660
```

```
QY      661 GGGCTCAAGTGATCCTCCTGCTCAGCTCCCAAGTAGCTGGGACTACAAACACAGTGTC 720
DB      661 GGGCTCAAGTGATCCTCCTGCTCAGCTCCCAAGTAGCTGGGACTACAAACACAGTGTC 720

QY      721 CCATGCTGGCTGATATTTTTTTTCTTGAACAGGGTATCAGCTCTGTGCCCCAGGCTGGA 780
DB      721 CCATGCTGGCTGATATTTTTTTTCTTGAACAGGGTATCAGCTCTGTGCCCCAGGCTGGA 780

QY      781 GTACAGTGGGTAATAATAGCTCAGCTCCCTCCTGCTGCTCAAGCAATCCGCTG 840
DB      781 GTACAGTGGGTAATAATAGCTCAGCTCCCTCCTGCTGCTCAAGCAATCCGCTG 840

QY      841 GCCTCAGCATCCTGAGTACAGGCTTGTGCCACAGGCCAGCTTAAGTTTT 900
DB      841 GCCTCAGCATCCTGAGTACAGGCTTGTGCCACAGGCCAGCTTAAGTTTT 900

QY      901 AAAAAATGATTTTGGTATAGAGGAGTCTGTATGTGTCTCAGGCTGATTTTTATTG 960
DB      901 AAAAAATGATTTTGGTATAGAGGAGTCTGTATGTGTCTCAGGCTGATTTTTATTG 960

QY      961 TTGAGACAGGCTCAGTATGTTGCCATGATCCGCCACCTCCACTCCCAAGTGCTCA 1020
DB      961 TTGAGACAGGCTCAGTATGTTGCCATGATCCGCCACCTCCACTCCCAAGTGCTCA 1020

QY      1021 TCTTATCTGTTCATTAGTCAGTTGACAGACATTTAGGTTGTTTCCACTTTTGGACCATTA 1080
DB      1021 TCTTATCTGTTCATTAGTCAGTTGACAGACATTTAGGTTGTTTCCACTTTTGGACCATTA 1080

QY      1081 TGAATAATCTCCAGTGAATATTCATGATATGTTGTTGGGCAATGTTTCAATTTCT 1140
DB      1081 TGAATAATCTCCAGTGAATATTCATGATATGTTGTTGGGCAATGTTTCAATTTCT 1140

QY      1141 GTTGGTGTATATCTAGGAGTGAATGCTGGATCCGGTAAATATTTGACAGGAGAG 1200
DB      1141 GTTGGTGTATATCTAGGAGTGAATGCTGGATCCGGTAAATATTTGACAGGAGAG 1200

QY      1201 TTCAGGGGAAGAAAACTTTGGGAAAAATGAGCATGTTTGAATAATCAGCAAGAGTGCAGGG 1260
DB      1201 TTCAGGGGAAGAAAACTTTGGGAAAAATGAGCATGTTTGAATAATCAGCAAGAGTGCAGGG 1260

QY      1261 GTTTTTCGGAGTTTATTTATTTATTTCTGTGACAAATGTCAGTTTGTAGTAAGTACAAG 1320
DB      1261 GTTTTTCGGAGTTTATTTATTTATTTCTGTGACAAATGTCAGTTTGTAGTAAGTACAAG 1320

QY      1321 TTATATAAGTGAAGAAGTGAATTAAGGCTGGAATAGGGCTTCAAGTAAATCATGA 1380
DB      1321 TTATATAAGTGAAGAAGTGAATTAAGGCTGGAATAGGGCTTCAAGTAAATCATGA 1380

QY      1381 AGCAGCTTTGAATACCAAAATTAAGGAGCTTGGCTGTAACAAAAATATAAAAAATCACA 1440
DB      1381 AGCAGCTTTGAATACCAAAATTAAGGAGCTTGGCTGTAACAAAAATATAAAAAATCACA 1440

QY      1441 TTTTTCCTTTTTCCTTTTGAAGAAGTCTTCTCTTTCACCTGCTGGAGGAGTGGTG 1500
DB      1441 TTTTTCCTTTTTCCTTTTGAAGAAGTCTTCTCTTTCACCTGCTGGAGGAGTGGTG 1500

QY      1501 TGATCTCAGCTCACTGCAACTTTTCGCTCCCGGGTTCAAGCAATTTCTCTGCTTCAGCCT 1560
DB      1501 TGATCTCAGCTCACTGCAACTTTTCGCTCCCGGGTTCAAGCAATTTCTCTGCTTCAGCCT 1560

QY      1561 CCCAAGTAGCTGGGACTACAGGCACATTCACCACATCCAGCTGATTTTGTATTTTAG 1620
DB      1561 CCCAAGTAGCTGGGACTACAGGCACATTCACCACATCCAGCTGATTTTGTATTTTAG 1620

QY      1621 TAGAGATGGATTTCTACTTTTGTGGCAAGCTGCTCAAACTTTTGTGCTCATATTTG 1680
DB      1621 TAGAGATGGATTTCTACTTTTGTGGCAAGCTGCTCAAACTTTTGTGCTCATATTTG 1680

QY      1681 TTGTAATCTATTTCTCTTTTGTGAGTGGGCCCCAGACCAAAAAAATAAATCTTAG 1740
DB      1681 TTGTAATCTATTTCTCTTTTGTGAGTGGGCCCCAGACCAAAAAAATAAATCTTAG 1740

QY      1741 AATCCAAATCAGTGCTGTTGTTTGGACCATGTCATTTGAGAACCAAGTGTGACAGGGC 1800
```

Db	1741	AAATCCAAATCAGTGTGTGTGGTTGACCACTGTCACTTCAGAAACCAACAGTGTGCACAGGGC	1800
Qy	1801	CTCAGGAGTAGAGGTGATCTCTGCTCGAAGAGAAATAGAAATGAAATATATCTCCGGGGCC	1860
Db	1801	CTCAGGAGTAGAGGTGATCTCTGCTCGAAGAGAAATAGAAATGAAATATATCTCCGGGGCC	1860
Qy	1861	AGGCGTGTGGCTCATGCGCTGTAAATCCAGCAGCACTTTGGAGGCCAAGGCATGTGGATCAC	1920
Db	1861	AGGCGTGTGGCTCATGCGCTGTAAATCCAGCAGCACTTTGGAGGCCAAGGCATGTGGATCAC	1920
Qy	1921	CTGAGGTCAGAGGTTCACAAACCCAGCGCTGCCAACATGCTGAAACCCCGCTCTCTACTAAAA	1980
Db	1921	CTGAGGTCAGAGGTTCACAAACCCAGCGCTGCCAACATGCTGAAACCCCGCTCTCTACTAAAA	1980
Qy	1981	ATACAAAAAATAGCTAAGTGTGGTGGCGCATGCGTGAATCCACAGTACTTTGGAGGGT	2040
Db	1981	ATACAAAAAATAGCTAAGTGTGGTGGCGCATGCGTGAATCCACAGTACTTTGGAGGGT	2040
Qy	2041	GAGCGAGGAGAAATTCCTTGAACCCGGAGGCGAGAGTTGTCAGTGAACGCGAGATCAACACCA	2100
Db	2041	GAGCGAGGAGAAATTCCTTGAACCCGGAGGCGAGAGTTGTCAGTGAACGCGAGATCAACACCA	2100
Qy	2101	CTGCACCTCCAGCGCTGGGGGAGAGAGCGAGACTTCCTCTCAAAAAACAACAAAACAAGA	2160
Db	2101	CTGCACCTCCAGCGCTGGGGGAGAGAGCGAGACTTCCTCTCAAAAAACAACAAAACAAGA	2160
Qy	2161	ATTAAGCAAAATAGACATTCAGAGAGAAACCTGAAGGGGGTCAGACACCACTACAGATTTC	2220
Db	2161	ATTAAGCAAAATAGACATTCAGAGAGAAACCTGAAGGGGGTCAGACACCACTACAGATTTC	2220
Qy	2221	TGTGCCACATGCCAAGTACTTCTGTAGGCGATGACTGGATGAGCTGTCCACATCTGAAATCA	2280
Db	2221	TGTGCCACATGCCAAGTACTTCTGTAGGCGATGACTGGATGAGCTGTCCACATCTGAAATCA	2280
Qy	2281	TCAGAGTCTGTTTCAGAACTTTCACACCGGACAGGAGCCAGGACTGGAATGCGAGTCTCCT	2340
Db	2281	TCAGAGTCTGTTTCAGAACTTTCACACCGGACAGGAGCCAGGACTGGAATGCGAGTCTCCT	2340
Qy	2341	GGTCACCTGGCCAGAGAGTTGGCGCTTGACCTGTAGACCACTGGCCCAACAAAGGAGCTGCTT	2400
Db	2341	GGTCACCTGGCCAGAGAGTTGGCGCTTGACCTGTAGACCACTGGCCCAACAAAGGAGCTGCTT	2400
Qy	2401	AGTCTACCTCCCAGGAAATCCACAGTGTCTTCTTCTGGGAAGTGAATCATTTGGCGCAG	2460
Db	2401	AGTCTACCTCCCAGGAAATCCACAGTGTCTTCTTCTTGGGAAGTGAATCATTTGGCGCAG	2460
Qy	2461	CATCTCGGTATTTTCTCCTCTTCCCAGGGGAAGATTCCTTAGGGCAGTATTTGGGAAAGACA	2520
Db	2461	CATCTCGGTATTTTCTCCTCTTCCCAGGGGAAGATTCCTTAGGGCAGTATTTGGGAAAGACA	2520
Qy	2521	TGGGCATGGAAGACACACCGGGTGAATGTCATACGCTTGCGTGTCTGAGCTCTCATGTTAA	2580
Db	2521	TGGGCATGGAAGACACACCGGGTGAATGTCATACGCTTGCGTGTCTGAGCTCTCATGTTAA	2580
Qy	2581	GGTCTCTACAGACAGGAAACATGGGGCACAGGACAGATCAGTAGGCTCAGAGCATC	2640
Db	2581	GGTCTCTACAGACAGGAAACATGGGGCACAGGACAGATCAGTAGGCTCAGAGCATC	2640
Qy	2641	TCAGGGACCGAGGGCAATATGTTCTCTGAGCGGGAATTAAGAGCTTGGGGCTCTCATATGGT	2700
Db	2641	TCAGGGACCGAGGGCAATATGTTCTCTGAGCGGGAATTAAGAGCTTGGGGCTCTCATATGGT	2700
Qy	2701	GTTTCTGGGCTCAACTGCCAGCTCCGTCACITACTTGTTGCTGTGACCATGGCGCAGTTA	2760
Db	2701	GTTTCTGGGCTCAACTGCCAGCTCCGTCACITACTTGTTGCTGTGACCATGGCGCAGTTA	2760
Qy	2761	TTCCATCTCTCCATATCTCTTTCCCTCACTTTTAAATGGAATATGGGTACCACTCC	2820
Db	2761	TTCCATCTCTCCATATCTCTTTCCCTCACTTTTAAATGGAATATGGGTACCACTCC	2820
Qy	2821	CAGGGTCACAGAGGCGTTACAGAAACGATCTTGTGAATTTGGCTTGCAGTAATAATTC	2880
Db	2821	CAGGGTCACAGAGGCGTTACAGAAACGATCTTGTGAATTTGGCTTGCAGTAATAATTC	2880

Db	2821	CAGGGTCACAGAGAGGCTTACAGAAACGATTCTTGTGAATTGGCTTCGACATAATAATTC	2888
Qy	2881	AAATACCTGCCAGCTATTCTTATTCCACATCCAAAGCCCTTTTCGCTCTGCTGCTGGGTGAAAA	2940
Db	2881	AAATACCTGCCAGCTATTCTTATTCCACATCCAAAGCCCTTTTCGCTCTGCTGCTGGGTGAAAA	2940
Qy	2941	CACATGTCAGTGTTTCCGTGACGGTTTCCACAAGAAGATTTCAAAAATTACAACTGCCAG	3000
Db	2941	CACATGTCAGTGTTTCCGTGACGGTTTCCACAAGAAGATTTCAAAAATTACAACTGCCAG	3000
Qy	3001	TCTGAAGAATTCCTCAAAAACATCCGCACGCATCTCTGGAGGCGCGGGCTTGGGGATGGGAC	3060
Db	3001	TCTGAAGAATTCCTCAAAAACATCCGCACGCATCTCTGGAGGCGCGGGCTTGGGGATGGGAC	3060
Qy	3061	TGCCCCCGGGTCTCGAACAGATGCTGCGCGCAGAGCACACACACACAGCCAGCCTG	3120
Db	3061	TGCCCCCGGGTCTCGAACAGATGCTGCGCGCAGAGCACACACACACAGCCAGCCTG	3120
Qy	3121	TGTGTGCGGCGGAGTCCGGTTCGCGGTGAGCAGCGCTGGCTGGTGGCGGGGCG	3180
Db	3121	TGTGTGCGGCGGAGTCCGGTTCGCGGTGAGCAGCGCTGGCTGGTGGCGGGGCG	3180
Qy	3181	AGAGCCATTGTTCCGABGCGTACCGAGAGCCCCCGCGCTCGCCCCGGGAGGAGCGGGGCT	3240
Db	3181	AGAGCCATTGTTCCGABGCGTACCGAGAGCCCCCGCGCTCGCCCCGGGAGGAGCGGGGCT	3240
Qy	3241	TCCGGGTCCCAAGCTCCAGATCTCGGGTGGCTGCCACCTCTCCCTGCCACGCGCTG	3300
Db	3241	TCCGGGTCCCAAGCTCCAGATCTCGGGTGGCTGCCACCTCTCCCTGCCACGCGCTG	3300
Qy	3301	GGGGAGCGGAAGACAGGGACGGAGATGTTAGTGTGGCGCCCCCGAGGGTTCAACACT	3360
Db	3301	GGGGAGCGGAAGACAGGGACGGAGATGTTAGTGTGGCGCCCCCGAGGGTTCAACACT	3360
Qy	3361	GTATCTGAGAACTTCCCGAGTCCACCCACCCGTTCTCCGTGTGCCCGAGGGCGGT	3420
Db	3361	GTATCTGAGAACTTCCCGAGTCCACCCACCCGTTCTCCGTGTGCCCGAGGGCGGT	3420
Qy	3421	CCTGGGCTAGGCTTCGCGCCCCAGCCCCAAACCGGGTCCCGAGCCCTTCCAGAGAGAAA	3480
Db	3421	CCTGGGCTAGGCTTCGCGCCCCAGCCCCAAACCGGGTCCCGAGCCCTTCCAGAGAGAAA	3480
Qy	3481	GCTCCCGACCGGGATTCGCGGCAGAGGCCACAGCGCGGGTGGAGAGAGCTGAGAAGG	3540
Db	3481	GCTCCCGACCGGGATTCGCGGCAGAGGCCACAGCGCGGGTGGAGAGAGCTGAGAAGG	3540
Qy	3541	AGAAACAGAGGGAGGGGAGCGAGAGCTGCGGCGCAGAGGAACACGATATCGCGCGA	3600
Db	3541	AGAAACAGAGGGAGGGGAGCGAGAGCTGCGGCGCAGAGGAACACGATATCGCGCGA	3600
Qy	3601	GCCAAATGGCAAGCGCAGGAGTGGCACCAAAATTCCTTCGCGCCAAATCACAGCCGGA	3660
Db	3601	GCCAAATGGCAAGCGCAGGAGTGGCACCAAAATTCCTTCGCGCCAAATCACAGCCGGA	3660
Qy	3661	GTTTACAGAGCCTCATTAGCATTTCCCCAGAGGCAGGGCAGAGGCGCGGGTGG	3720
Db	3661	GTTTACAGAGCCTCATTAGCATTTCCCCAGAGGCAGGGCAGAGGCGCGGGTGG	3720
Qy	3721	TGTGGTTCGGTTCGGCAGCATCCCGCGCGCCTGCTGCGGTTCGCGCGAGCCTCGGCC	3780
Db	3721	TGTGGTTCGGTTCGGCAGCATCCCGCGCGCCTGCTGCGGTTCGCGCGAGCCTCGGCC	3780
Qy	3781	TCTGTCTCTCCCTCCCGCCCTTACCTCCAGCGGGGACCGCCCGCGCCAGTCAACTCC	3840
Db	3781	TCTGTCTCTCCCTCCCGCCCTTACCTCCAGCGGGGACCGCCCGCGCCAGTCAACTCC	3840
Qy	3841	TCGCACTTGGCCCTCTTGGCAGCGGATAAAGGGGGCTGAGGAATACCGGACACGGT	3900
Db	3841	TCGCACTTGGCCCTCTTGGCAGCGGATAAAGGGGGCTGAGGAATACCGGACACGGT	3900
Qy	3901	CACCCGTGCCAGCTCTAGCCCTTTAAATTCCTGGCTCGGGGACCTCACACACCGCGGT	3960
Db	3901	CACCCGTGCCAGCTCTAGCCCTTTAAATTCCTGGCTCGGGGACCTCACACACCGCGGT	3960

	CC	complications. (I) or (II) is useful as research agent and materials for
	CC	discovery of treatments and diagnostics for a disease, particularly human
	CC	disease. (II) is useful for constructing nucleotide probes and primers, of
	CC	for detecting genetic polymorphism, for detecting changes in the level of
	CC	expression of (II), and as a diagnostic tool. This sequence represents
	CC	a control region from a gene encoding a fatty acid regulated protein.
XX		
SQ		Sequence 1970 BP; 428 A; 590 C; 588 G; 364 T; 0 other;
		Query Match            46.6%; Score 1932.8; DB 24; Length 1970;
		Best Local Similarity 99.7%; Pred. No. 0;
		Matches 1967; Conservative         0; Mismatches         2; Indels         3; Gaps         3
QY	2121	GAGCGGAGACTTCTCTCAAAAAACAAAACAAGAATTAAGCAAATTAGCAATTG 2180
DB	1	GAGCGGAGACTTCTCTCAAAAAACAAAACAAGAATTAAGCAAATTAGCAATTG 60
QY	2181	CAGAGAACTGAAGGGGTGCAGACCAGTCACAGATTCTCTGCCACATGCCAAGTACT 2240
DB	61	CAGAGAACTGAAGGGGTGCAGACCAGTCACAGATTCTCTGCCACATGCCAAGTACT 120
QY	2241	TCTGAGGCATGACTGGATGAGTGCTCCTCAATCATCCAGTCTTTGTCAGAACTT 2300
DB	121	TCTGAGGCATGACTGGATGAGTGCTCCTCAATCATCCAGTCTTTGTCAGAACTT 180
QY	2301	TCACACCGCAGGGAGCCAGACTGGAAATGACAGTCTCTGTGCTACTGGCCAGAGTTG 2360
DB	181	TCACACCGCAGGGAGCCAGACTGGAAATGACAGTCTCTGTGCTACTGGCCAGAGTTG 240
QY	2361	GCCTTGACCCCTGAGACCAAGTGCCCAACAAAGGAGTGCCTTAGCTACCTCCCAGAAATC 2420
DB	241	GCC-TGACCCCTGAGACCAAGTGCCCAACAAAGGAGTGCCTTAGCTACCTCCCAGAAATC 299
QY	2421	CCAGSTGCTTGTCTTCTCTGGGAAGTGAATCATTTGGCGCAGCACTCCGTATTTCTCCTCT 2480
DB	300	CCAGSTGCTTGTCTTCTCTGGGAAGTGAATCATTTGGCGCAGCACTCCGTATTTCTCCTCT 359
QY	2481	TCCAGGGGAAGGATCTTAGGCCAGTATTTGGGAAGACATGGGCAATGGAAGCACACCG 2540
DB	360	TCCAGGGGAAGGATCTTAGGCCAGTATTTGGGAAGACATGGGCAATGGAAGCACACCG 419
QY	2541	GTGAATGCATAGCCGCTGGTTCGAGCTCTCATGGTAAGGCTCTACAGACACCGAAA 2600
DB	420	GTGAATGCATAGCCGCTGGTTCGAGCTCTCATGGTAAGGCTCTACAGACACCGAAA 479
QY	2601	AGATGGGGGCACAGGCACAGATFCAGTAGGGTCAGAGCAFTCTCAGGGACCGAGGGCAATAF 2660
DB	480	AGATGGGGGCACAGGCACAGATFCAGTAGGGTCAGAGCAFTCTCAGGGACCGAGGGCAATAF 539
QY	2661	GGTCTGAGCAGGGATTAAGAGCTTGGGCTCTCATATGGTGTTTTCTGGGCTCAACTGCCA 2720
DB	540	GGTCTGAGCAGGGATTAAGAGCTTGGGCTCTCATATGGTGTTTTCTGGGCTCAACTGCCA 599
QY	2721	GCTCGGTCACTATACRTGGTTCGTGACCATGGCCAAGTTATTCCCATCTCTCCATATCTCT 2780
DB	600	GCTCCATCACTTACTTGTTGCTGTGACCATGGCAAGTTATTCCCATCTCTCCATATCTCT 659
QY	2781	TTCTCTACTTTTAAATGGAATAATGGGGTACCCACCTCCAGGGTTCACAGAGAGCTTA 2840
DB	660	TTCTCTACTTTTAAATGGAATAATGGGGTACCCACCTCCAGGGTTCACAGAGAGCTTA 719
QY	2841	CAGAAACGATCTTGTAATGGCTTCGAGTAATAATTCAATACCTGCCAGCTATCTT 2900
DB	720	CAGAAACGATCTTGTAATGGCTTCGAGTAATAATTCAATACCTGCCAGCTATCTT 779
QY	2901	ATTCCACATCCAAGCCCTTTGCCCTGCCTGGGTGAAACACATGTCAGTGTTCCTCGA 2960
DB	780	ATTCCACATCCAAGCCCTTTGCCCTGCCTGGGTGAAACACATGTCAGTGTTCCTCGA 839
QY	2961	CGGTTTCCAAAAGAAGATTC AAAATATACAACTGCCAGTCTGAAGAAATCTCCAAAACA 3020
DB	840	CGGTTTCCAAAAGAAGATTC AAAATATACAACTGCCAGTCTGAAGAAATCTCCAAAACA 899







[illegible]

```
Db 17404 AACCGCAGTCTCCGCGACCCGAACTCGCTCCGAGGCTCAGCCCTGGAAAGTGA 17345
QY 4082 TCCCGGCATCCGAGACCCAGATGCGCGGCCACTTCTCTCAGGACGATGTGAGTTTCCCA 4141
Db 17344 TCCCGGCATCCGAGACCCAGATGCGCGGCCACTTCTCTCAGGACGATGTGAGTTTCCCA 17285
QY 4142 GCCTGSCCC 4150
Db 17284 GCCTGGCCC 17276

RESULT 5
ID ABK42977/c
XX ID ABK42977 standard; DNA; 17707 BP.
AC ABK42977;
XX
DT 21-MAY-2002 (first entry)
XX
DE Genomic sequence #876 encoding novel human connective tissue polypeptide.
XX
KW Human; connective tissue related disorder; cancer; gene therapy;
KW cytosstatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200155343-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01322.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
```



PR 17-NOV-2000; 2000US-0249213.





Db 56634 TAGAGACGGGTTTCACCTGTGTGATGCCAGGATGGTCTGGGACTCCTGACCTTGTGATCGG 56575  
Qy 1670 ---TGTCAATAATGTTGTTAACTATTGTTCCTTTTGCTGAGGTAGGGCCGCCAGACCCAAA 1725  
Db 56574 CCGGCTCGACCTCTCAAAAGTCTGGGATTACAGCGGTGAGCCACCGGACCGCGGAGTTT 56515  
Qy 1726 AAAATAAATCTTGAATCCAAATCAGTGTGTGTTTGACACACTGTCTACCTTTGAGAACC 1785  
Db 56514 ATTTGGAATCTCTTCATCGATGGAGATTGGCCTATTCTCCCTATTATTTCATTCATCCAT 56455  
Qy 1786 CAGTGTGACCGAGGCTCAGGAGTATAGGTATCTCTGCTGAAAGAGAAATAGATGAA 1845  
Db 56454 TTATATCAGCATGGACTTACAGGTATTTATTTATATCTTTGGATATGATCTCAAACTAC 56395  
Qy 1846 AATATTTCTCC-----GGGCCAGGCGTGTGGCTCATGCTCTTAATCCAGCACCTTT 1896  
Db 56394 TTAATTTGTTGCCAAGTGTGCCAGGCGTGTGGCTCACCTCTAGTCCAGCACCTTT 56335  
Qy 1897 GGGAGCCCAAGCAGTGTGGATCACCTGAGGTCAGGAGTTCAAAACCAGCTTGGCCCAACAT 1956  
Db 56334 GGGAGCCCGAGCAGCGGATCAC--GAGTCAAGGATCAAGACCATCTTGGCTTAACAC 56277  
Qy 1957 GGTGAACCCCGTCTCTACTAAAAATACAAAAATTAGTATAGTGTGGTGGCGATGCGCT 2016  
Db 56276 AGTGAACCTTGTCTCTATTAAAAATCAAAAAATTAGCCAGGCATGGTGGCGGCGCT 56217  
Qy 2017 GTATCCAGCTACTTGGGAGGTTGAGCAGGAGAAATTTCTTGAACCGGAGGCGAGG 2076  
Db 56216 GTAGTCCAGCTACTTGGAGGCTGAGCAGGAGAAATGTTGAACCCGAGGCGGAGC 56157  
Qy 2077 TTGAGTGAACGGAGATCACACCTGTCACCTGACCTGCGGAGGAGAGAGAGCTTCCT 2136  
Db 56156 TTGCAGTGAGCGGAGATCGCCCACTGC-CTCCAGCCTGGGTGACAGAGCGAGACTCCAT 56098  
Qy 2137 CTCAAAAAACAACAAAAACAAAGATTT 2163  
Db 56097 CTCAAAAAGTAAATAAAAAATTT 56071  
  
RESULT 8  
AAK69044/c  
ID AAK69044 standard; DNA; 5491 BP.  
XX  
AC AAK69044;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23856.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytosolic; gene therapy; vaccine; metastasis; ds.  
XX OS Homo sapiens.  
XX PN W0200157182-A2.  
XX PD 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184654.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216860.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217490.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226668.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
XX 14-SEP-2000; 2000US-0232399.  
XX 14-SEP-2000; 2000US-0232400.  
XX 14-SEP-2000; 2000US-0232401.  
XX 14-SEP-2000; 2000US-0233063.  
XX 14-SEP-2000; 2000US-0233064.  
XX 14-SEP-2000; 2000US-0233065.  
XX 21-SEP-2000; 2000US-0234223.  
XX 21-SEP-2000; 2000US-0234274.  
XX 25-SEP-2000; 2000US-0234997.  
XX 25-SEP-2000; 2000US-0234998.  
XX 26-SEP-2000; 2000US-0235484.  
XX 27-SEP-2000; 2000US-0235834.  
XX 27-SEP-2000; 2000US-0235836.  
XX 29-SEP-2000; 2000US-0236327.  
XX 29-SEP-2000; 2000US-0236367.  
XX 29-SEP-2000; 2000US-0236368.  
XX 29-SEP-2000; 2000US-0236369.  
XX 29-SEP-2000; 2000US-0236370.  
XX 02-OCT-2000; 2000US-0236802.  
XX 02-OCT-2000; 2000US-0237037.  
XX 02-OCT-2000; 2000US-0237038.  
XX 02-OCT-2000; 2000US-0237039.  
XX 02-OCT-2000; 2000US-0237040.  
XX 13-OCT-2000; 2000US-0239935.  
XX 13-OCT-2000; 2000US-0239937.  
XX 20-OCT-2000; 2000US-0240960.  
XX 20-OCT-2000; 2000US-0241221.  
XX 20-OCT-2000; 2000US-0241785.  
XX 20-OCT-2000; 2000US-0241786.  
XX 20-OCT-2000; 2000US-0241787.  
XX 20-OCT-2000; 2000US-0241808.  
XX 20-OCT-2000; 2000US-0241809.  
XX 20-OCT-2000; 2000US-0241809.







PR	14-SEP-2000;	2000US-02342000;	2000US-02342000;
PR	14-SEP-2000;	2000US-02342001;	2000US-02342001;
PR	14-SEP-2000;	2000US-02343063;	2000US-02343063;
PR	14-SEP-2000;	2000US-02343064;	2000US-02343064;
PR	14-SEP-2000;	2000US-02343065;	2000US-02343065;
PR	21-SEP-2000;	2000US-02342223;	2000US-02342223;
PR	21-SEP-2000;	2000US-02342273;	2000US-02342273;
PR	25-SEP-2000;	2000US-02349977;	2000US-02349977;
PR	25-SEP-2000;	2000US-02349998;	2000US-02349998;
PR	26-SEP-2000;	2000US-02344884;	2000US-02344884;
PR	27-SEP-2000;	2000US-02358634;	2000US-02358634;
PR	27-SEP-2000;	2000US-02358636;	2000US-02358636;
PR	29-SEP-2000;	2000US-02365377;	2000US-02365377;
PR	29-SEP-2000;	2000US-02365367;	2000US-02365367;
PR	29-SEP-2000;	2000US-02363687;	2000US-02363687;
PR	29-SEP-2000;	2000US-02363669;	2000US-02363669;
PR	29-SEP-2000;	2000US-02363670;	2000US-02363670;
PR	13-OCT-2000;	2000US-02399357;	2000US-02399357;
PR	13-OCT-2000;	2000US-02399367;	2000US-02399367;
PR	20-OCT-2000;	2000US-02411821;	2000US-02411821;
PR	20-OCT-2000;	2000US-02411826;	2000US-02411826;
PR	20-OCT-2000;	2000US-02417856;	2000US-02417856;
PR	20-OCT-2000;	2000US-02417877;	2000US-02417877;
PR	20-OCT-2000;	2000US-02418088;	2000US-02418088;
PR	20-OCT-2000;	2000US-02418109;	2000US-02418109;
PR	20-OCT-2000;	2000US-02418266;	2000US-02418266;
PR	01-NOV-2000;	2000US-02446177;	2000US-02446177;
PR	08-NOV-2000;	2000US-02464744;	2000US-02464744;
PR	08-NOV-2000;	2000US-02464755;	2000US-02464755;
PR	08-NOV-2000;	2000US-02465233;	2000US-02465233;
PR	08-NOV-2000;	2000US-02465224;	2000US-02465224;
PR	08-NOV-2000;	2000US-02465255;	2000US-02465255;
PR	08-NOV-2000;	2000US-02465266;	2000US-02465266;
PR	08-NOV-2000;	2000US-02465277;	2000US-02465277;
PR	08-NOV-2000;	2000US-02465288;	2000US-02465288;
PR	08-NOV-2000;	2000US-02465332;	2000US-02465332;
PR	08-NOV-2000;	2000US-02466609;	2000US-02466609;
PR	08-NOV-2000;	2000US-02466110;	2000US-02466110;
PR	08-NOV-2000;	2000US-02466133;	2000US-02466133;
PR	17-NOV-2000;	2000US-02492077;	2000US-02492077;
PR	17-NOV-2000;	2000US-02492088;	2000US-02492088;
PR	17-NOV-2000;	2000US-02492099;	2000US-02492099;
PR	17-NOV-2000;	2000US-02492111;	2000US-02492111;
PR	17-NOV-2000;	2000US-02492133;	2000US-02492133;
PR	17-NOV-2000;	2000US-02492144;	2000US-02492144;
PR	17-NOV-2000;	2000US-02492155;	2000US-02492155;
PR	17-NOV-2000;	2000US-02492166;	2000US-02492166;
PR	17-NOV-2000;	2000US-02492177;	2000US-02492177;
PR	17-NOV-2000;	2000US-02492188;	2000US-02492188;
PR	01-DEC-2000;	2000US-02501600;	2000US-02501600;
PR	01-DEC-2000;	2000US-02503991;	2000US-02503991;
PR	05-DEC-2000;	2000US-02510390;	2000US-02510390;
PR	05-DEC-2000;	2000US-02519888;	2000US-02519888;
PR	05-DEC-2000;	2000US-02567199;	2000US-02567199;

PR	06-DEC-2000; 2000US-0251479.	
PR	08-DEC-2000; 2000US-0251856.	
PR	08-DEC-2000; 2000US-0251856.	
PR	08-DEC-2000; 2000US-0251868.	
PR	08-DEC-2000; 2000US-0251869.	
PR	08-DEC-2000; 2000US-0251889.	
PR	08-DEC-2000; 2000US-0251890.	
PR	11-DEC-2000; 2000US-0254597.	
PR	05-JAN-2001; 2001US-0259678.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-465566/50.	
XX		
PT	Novel polypeptides and polynucleotides useful for diagnosing,	
PT	preventing, treating neural, immune system, muscular, reproductive,	
PT	pulmonary, cardiovascular, renal, proliferative disorders and cancerous	
PT	diseases	
XX		
PS	Disclosure; SEQ ID No 2141; 1180pp; English.	
XX		
CC	The present invention relates to the isolation of novel human enzyme	
CC	polypeptides (AAU2915-AAU23814), and the cDNA and genomic sequences	
CC	encoding them. The enzyme polypeptides of the invention may comprise the	
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,	
CC	isomerases or ligases. The sequences of the invention are useful in the	
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of	
CC	disorders including hyperproliferative disorders (e.g. cancer),	
CC	immunodeficiency disorders (e.g. AIDS) autoimmune disorders	
CC	(e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),	
CC	metabolic disorders (e.g. phenylketonuria), inflammatory disorders	
CC	(e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),	
CC	blood-related disorders (e.g. haemophilia), reproductive disorders	
CC	(e.g. infertility) and infectious disorders (e.g. influenza). The	
CC	polynucleotides of the invention can also be used in gene therapy.	
CC	AA541685-AA542192 represent DNA sequences encoding for the novel human	
CC	enzyme polypeptides of the invention.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 9742 BP; 2382 A; 2564 G; 2516 G; 2280 T; 0 other;	
	Query Match 8.8%; Score 366.4; DB 22; Length 9742;	
	Best Local Similarity 56.3%; Pred. No. 2.4e-64;	
	Matches 1043; Conservative 0; Mismatches 631; Indels 178; Gaps 1	
Qy	370 TATTATTAGACAGGTTCTCACTC-TGTCACCCAGCGTGGAGTGCAGTGCACAAATCAT 428	
Db		
Db	6351 TTTTCTTTGATGGAGTTTCACTCTTTGTTGCCAGCGTGGAGTGTGCTGGCGCTATCTT 6292	
Qy	429 AGCTCATGTAACTCAACTCTCGGCTCAAGTGATCTTACTTACCTCAGCTCCGAGGT 488	
Db		
Db	6291 GGCTCATGTCAACCTCGCGCTCCGGGTTCAAGCAATTCCTCGCTCAGCCTCCGAGT 6232	
Qy	489 AGCTAGGACTACAGGCACACACAGCCATCTAGTGGCTAAATTTTTTTTAAATTTTCATTT 548	
Db		
Db	6231 AGCTGGGTTACAGGCATGTGCCACACGCCCGCGCTAAATTTTGTATTTTA----- 6181	
Qy	549 TATGTATTCAPTTTCTTTCTTTTGTGTTGTTTGTGATATAGGGTCTCACTTTGTT 608	
Db		
Db	6180 -----GTAGCAGCAGGGTTTCTTCCATGTT 6157	
Qy	609 ACCCAGCGTGGAGGCGAGTGGCATGGTGACAGTGACGACCCCTTGACTTCTCTGGGCTCAA 668	
Db		
Db	6156 GGTCAAGCTGGT-----CTTGAACCTCCGACCTCAG 6126	
Qy	669 GTGATCTCTGCTCAGCGCTCCCAAGTAGCTGGGACTACAAACAGTGTCAACATGGCT 728	
Db		
Db	6125 GTGATCCGCCGCGCTCCGCTTCCCAAAGTGTGGGATTAACAGCGGTGACCCACAGCGCTC 6066	
Qy	729 GGCTGATATTTTTTCTTGAAACAGGGTATCATCTGTGTGCCAGGCTGGAGTACAGTG 788	













QY	1621	TAGAGATGGGATTTCACCTTTGTTGGCCAAAGCTGGTCTCAAACACTTTTTCCTGTCTATAATTG	1680	07-JUN-2000; 2000US-0209457
Db	3805	TACAGATGGGGTTTCACCATGTGGTCAAGCTGGTCTCGAAGCTCCTGACTTCATGATCCG	3864	28-JUN-2000; 2000US-0214886
QY	1681	TTGTAACTATTGTTCCTTTTTCCTGAGGTAGGGCCCCAGACCCAAACAAAATAATCTTAG	1740	PR 30-JUN-2000; 2000US-0215135
Db	3865	TCGCCCTCGACCTCCCAAGTGTTAGGATTACAGGCATGAGCCACTGTGCCAGCC----	3920	PR 07-JUL-2000; 2000US-0216647
QY	1741	AATCCAAATCAGTGTGTGTTGTTGACCACTGTCTACTTGAAGAACACACAGTGTGACAGGGC	1800	PR 07-JUL-2000; 2000US-0216880
Db	3921	---GCTAATTATTTTGTACAGACAGAGTTTCTCCATGTTGCCAGGCTGATCTGGAA	3977	PR 11-JUL-2000; 2000US-0217487
QY	1801	CTCAGGAGTATAGGTGATCTCTGCTCGAAGAGAAATAGAAATATATCTCCGGGCC	1860	PR 11-JUL-2000; 2000US-0217496
Db	3978	CTCTGGGGCCAAAGTGAACCATCTGCCTCGGC---CTCCAAGTCTGGGATTACAGGCC	4034	PR 14-JUL-2000; 2000US-0220963
QY	1861	AGGCGTGTGCTCATCGCTGTATCTCCAGCACACTTTTGGGAGGCCAAGGCATGTGGATCAC	1920	PR 26-JUL-2000; 2000US-0220963
Db	4035	GGGTGCGGTGGCTCACGCCCTGTATATTCAGCACCTTTGGGAGGCCAAGCGAGTGGATCAG	4094	PR 14-AUG-2000; 2000US-0224518
QY	1921	CTGAGTTCAGGAGTTCAAAACAGCCCTGGCCAAACATGCTGTAACCCCGTCTCTACTAAAA	1980	PR 14-AUG-2000; 2000US-0225213
Db	4095	CTGAGTTCGAGGTTCAAGACCAGCCCTGGCCAAACAGGTGACACCCCTGTCTACTAAAA	4154	PR 14-AUG-2000; 2000US-0225214
QY	1981	ATACAAAAATTAGCTAAGTGTGTGGCGATGCTGTATCCAGCTACTCTGTTGGAGGGT	2040	PR 14-AUG-2000; 2000US-0225267
Db	4155	ATAC-AAAAATTAGCCAGTGTGGTGGCGCATGCTATAGTCCCACTACTCTGAGAGGCT	4213	PR 14-AUG-2000; 2000US-0225268
QY	2041	GAGCAGAGGAAATTTCTGACCCGGGAGGCAGAGTTCAGTGAAGGAGATCACACCA	2100	PR 14-AUG-2000; 2000US-0225447
Db	4214	GAGCAGGAGAAATCATTGAACCCAGGAGGGAGTTGCAAGTGAAGGAGATCACACCA	4273	PR 14-AUG-2000; 2000US-0225758
QY	2101	CTGCACCTCCAGCCCTGGGGAGAGAGCGAGACTCTCTCAAAAAACAAAAACAAAAAGA	2160	PR 14-AUG-2000; 2000US-0225759
Db	4274	CTGCACCTCCAGCCCTGGGGAGAGAGTGAAGTGGTGTCTCTCAAAACAAACAAACAAA	4333	PR 16-AUG-2000; 2000US-0226279
QY	2161	ATTAAGCAAAATTAGACA	2177	PR 22-AUG-2000; 2000US-0226681
Db	4334	ACAAAACAAACAAACAA	4350	PR 22-AUG-2000; 2000US-0226688
RESULT	14			PR 22-AUG-2000; 2000US-0227182
ABL98014	ID	ABL98014 standard; DNA; 13216 BP.		PR 23-AUG-2000; 2000US-0227009
XX	XX			PR 30-SEP-2000; 2000US-0228924
AC	AC			PR 01-SEP-2000; 2000US-0229287
XX	XX			PR 01-SEP-2000; 2000US-0229343
DT	DT	21-JUN-2002 (first entry)		PR 01-SEP-2000; 2000US-0229344
XX	XX	Human testicular antigen encoding DNA fragment SEQ ID NO: 2666.		PR 05-SEP-2000; 2000US-0229509
DE	DE	Human; testicular antigen; testes; cancer; metastasis; immune disorder;		PR 05-SEP-2000; 2000US-0229513
XX	XX	reproductive system disorder; urinary system disorder; gene therapy;		PR 06-SEP-2000; 2000US-0230437
KW	KW	cardiovascular disorder; respiratory disorder; neurological disorder;		PR 06-SEP-2000; 2000US-0230438
KW	KW	gastrointestinal disease; infection; cytostatic; gene; ds.		PR 08-SEP-2000; 2000US-0231243
XX	XX	Homo sapiens.		PR 08-SEP-2000; 2000US-0231244
XX	XX	WO200155317-A2.		PR 08-SEP-2000; 2000US-0231413
PD	PD	02-AUG-2001.		PR 08-SEP-2000; 2000US-0231414
XX	XX			PR 08-SEP-2000; 2000US-0232080
XX	XX			PR 08-SEP-2000; 2000US-0232081
XX	XX			PR 12-SEP-2000; 2000US-0231968
XX	XX			PR 14-SEP-2000; 2000US-0232397
XX	XX			PR 14-SEP-2000; 2000US-0232398
XX	XX			PR 14-SEP-2000; 2000US-0232399
XX	XX			PR 14-SEP-2000; 2000US-0232400
XX	XX			PR 14-SEP-2000; 2000US-0232401
XX	XX			PR 14-SEP-2000; 2000US-0233063
XX	XX			PR 14-SEP-2000; 2000US-0233064
XX	XX			PR 14-SEP-2000; 2000US-0233065
XX	XX			PR 21-SEP-2000; 2000US-0234223
XX	XX			PR 21-SEP-2000; 2000US-0234274
XX	XX			PR 25-SEP-2000; 2000US-0234997
XX	XX			PR 25-SEP-2000; 2000US-0234998
XX	XX			PR 26-SEP-2000; 2000US-0235484
XX	XX			PR 27-SEP-2000; 2000US-0235834
XX	XX			PR 27-SEP-2000; 2000US-0235836
XX	XX			PR 29-SEP-2000; 2000US-0236327
XX	XX			PR 29-SEP-2000; 2000US-0236367
XX	XX			PR 29-SEP-2000; 2000US-0236368
XX	XX			PR 29-SEP-2000; 2000US-0236369
XX	XX			PR 29-SEP-2000; 2000US-0236370
XX	XX			PR 02-OCT-2000; 2000US-0236802
XX	XX			PR 02-OCT



Search completed: May 20, 2003, 14:53:49  
Job time : 1583 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 14:24:42 ; Search time 134 seconds  
(without alignments)  
9497.825 Million cell updates/sec

Title: US-10-016-725-15

Perfect score: 4150

Sequence: 1 aggaattaccatttaaat.....tgagttccagcctggccc 4150

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	349.8	8.4	12394	4	US-09-488-856A-10
2	345	8.3	84495	4	US-09-797-906-3
3	321	7.7	35060	3	US-08-814-095-7
4	316.6	7.6	59065	4	US-09-813-817-3
5	316.6	7.6	59065	4	US-09-978-197-3
6	308.2	7.4	162450	4	US-09-345-882-1
7	307.6	7.4	43950	4	US-09-735-934A-3
8	307.2	7.4	70000	4	US-09-851-896-3
9	303	7.3	43950	4	US-08-724-394A-20
10	301.8	7.3	246240	2	US-08-724-394A-21
11	301.8	7.3	246240	2	US-08-724-394A-22
12	297.2	7.2	70000	4	US-09-851-896-3
13	296.6	7.1	59065	4	US-09-813-817-3
14	296.6	7.1	59065	4	US-09-978-197-3
15	288	6.9	21234	3	US-09-810-671-3
16	286.8	6.9	35060	3	US-08-814-095-7
17	285.8	6.9	162450	4	US-09-345-882-1
18	283.8	6.8	12394	4	US-09-488-856A-10
19	282.8	6.8	246240	2	US-08-724-394A-20
20	282.8	6.8	246240	2	US-08-724-394A-21
21	282.8	6.8	246240	2	US-08-724-394A-22
22	275	6.6	84495	4	US-09-797-906-3
23	270.4	6.5	72604	4	US-09-268-992-7
24	270.4	6.5	72604	4	US-09-657-474-7
25	267.2	6.4	45546	4	US-09-146-053-6
26	263	6.3	98844	4	US-09-791-211-10
27					

C 28	262.8	6.3	29629	4	US-09-729-995-3	Sequence 3, Appli
C 29	261.2	6.3	17327	1	US-07-906-871-15	Sequence 15, Appli
C 30	260.6	6.3	7676	1	US-08-451-777A-7	Sequence 7, Appli
C 31	260.6	6.3	7676	2	US-08-451-778A-7	Sequence 7, Appli
C 32	260.6	6.3	7676	2	US-08-998-208-7	Sequence 7, Appli
C 33	260.6	6.3	7676	5	PCT-US95-06743-7	Sequence 7, Appli
C 34	259.4	6.3	17327	1	US-07-906-871-15	Sequence 15, Appli
C 35	258	6.2	152331	3	US-09-128-155-16	Sequence 16, Appli
C 36	258	6.2	152331	3	US-09-128-155-16	Sequence 16, Appli
C 37	258	6.2	176373	3	US-09-128-155-17	Sequence 17, Appli
C 38	257.2	6.2	29629	4	US-09-729-995-3	Sequence 3, Appli
C 39	256.8	6.2	4326	2	US-08-852-807-12	Sequence 12, Appli
C 40	256.8	6.2	13674	2	US-08-852-807-1	Sequence 1, Appli
C 41	254.8	6.1	176373	3	US-09-128-155-17	Sequence 17, Appli
C 42	248.8	6.0	99500	4	US-09-798-096-10	Sequence 10, Appli
C 43	244	5.9	4326	2	US-08-852-807-12	Sequence 12, Appli
C 44	244	5.9	13674	2	US-08-852-807-1	Sequence 1, Appli
C 45	243.4	5.9	99500	4	US-09-798-096-10	Sequence 10, Appli

## ALIGNMENTS

RESULT 1  
US-09-488-856A-10  
; Sequence 10, Application US/09488856A  
; Patent No. 6316259  
; GENERAL INFORMATION:  
; APPLICANT: Brett P. Monia  
; APPLICANT: Robert McKay  
; APPLICANT: Madeline M. Butler  
; APPLICANT: Jacqueline Wyatt  
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLYCOGEN SYNTHASE KINASE 3 ALPHA  
; FILE REFERENCE: RUS-0115  
; CURRENT APPLICATION NUMBER: US/09/488,856A  
; CURRENT FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 88  
; SEQ ID NO 10  
; LENGTH: 12394  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (115)...(397)  
; NAME/KEY: CDS  
; LOCATION: (2438)...(2625)  
; NAME/KEY: CDS  
; LOCATION: (5639)...(5722)  
; NAME/KEY: CDS  
; LOCATION: (5864)...(5974)  
; NAME/KEY: CDS  
; LOCATION: (7902)...(8032)  
; NAME/KEY: CDS  
; LOCATION: (8121)...(8227)  
; NAME/KEY: CDS  
; LOCATION: (9197)...(9294)  
; NAME/KEY: CDS  
; LOCATION: (9375)...(9470)  
; NAME/KEY: CDS  
; LOCATION: (9898)...(10084)  
; NAME/KEY: CDS  
; LOCATION: (10431)...(10523)  
; NAME/KEY: CDS  
; LOCATION: (11713)...(11786)  
US-09-488-856A-10

Query Match 8.4%; Score 349.8; DB 4; Length 12394;  
Best Local Similarity 55.2%; Pred. No. 9.6e-76;  
Matches 1016; Conservative 0; Mismatches 742; Indels 81; Gaps 14;

QY 370 TATTATTAGACAGGTTCTCCTGTCACCCAGGCTGGAGTGGCAGTCAATCAATA 429  
Db 3604 TTTTITGATGAGTCTTCTGTCACCCAGGCTGGAGTGGCAGTCTTG 3663

QY 430 GCTCACTGTAACCTCAAACTCTCGGGCTCAAGTGTACTACCTACCTCCAGAGTA 489  
DB 3664 GCCCGTACAACCTCTGCTCCAGGTTCAAGCGATTCTTGCTCCCAACCACTCAGTA 3723  
QY 490 GTTAGCATTACAGGACACACAGCCATACCTGGCTAAATTTTTTTTTTAATTTCAATTT 549  
DB 3724 GCTGGGATTGACAGGATGGCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3781  
QY 550 ATGTATTCATTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 609  
DB 3782 GTAGAGACAGGTTTGGCCATGTTGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 3841  
QY 610 CCCAGCTGGAGGCGAGTGGCATGTTGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 669  
DB 3842 CGCCCCCACCCTCGACCTCCCAAGTGTGGTATTACAGGCATGA-----GCCAGCAT 3894  
QY 670 TGATCCTCTGCTCAGCTCAGCTCCCAAGTGTGGTATTACAGGCATGA-----GCCAGCAT 3894  
DB 3895 ACCTAGCCCTGATTTTCAAGACAACACTGAAACTGGATTTAGATGTGAAATCTTTTTTT 3954  
QY 729 GCTGATATTTTTTTCTTTGAACAGGGTATCACTCTGTGCTGCTGCTGCTGCTGCTGCTGCT 788  
DB 3955 TTTTTTTTTTTTTTTTTTTTGTAGAGGAGTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 4014  
QY 789 GGTAAATAATAGCTCACTGCAGCTCC-CCCTCTGGGCTCAAGCAATCGCTGGCTCAG 847  
DB 4015 GCGTATCTCGGGTCACTGCAACCTCCCGCTGCGGGTTCAGCGATTTCTTCGCTCAG 4074  
QY 848 CATCTAGTAGCTGGGACTCAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 907  
DB 4075 CCTCCTAGTAGCTGGGACTCAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4134  
QY 908 GATTTTGGTATAGGAGGCTTCTGCTATGTTGCTCAGGCTGATTTTATTTGTTG----- 963  
DB 4135 CTAGTAGGAGGCGGGTTTCAACGAGTTAGCCAGGATGCTCTATTTTTTTTTTTTTTT 4194  
QY 964 ----AGACAAGTCTCACTATGTTGCCATGATCCCCCACTCCACTTCCCAAGTGTCTC 1019  
DB 4195 TTTAAGACAGAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4254  
QY 1020 ATCTATCTGTTATAGTCACTAGTGCACAGACATTTAGTGTGTTTCCACTTTTTCACCAAT 1079  
DB 4255 CTGCAACCTCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4314  
QY 1080 ATGAATAATCACTCAAGTAATATATATATATATATATATATATATATATATATATAT 1139  
DB 4315 TCTTGTACCTCATCCACTGAGTATGTTGAATACAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 4374  
QY 1140 TGTGGGTTTATATCTAGGAGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199  
DB 4375 ATTTTGTATTTTATAGTAGATGGGTTTGGCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 4434  
QY 1200 GTTCAGGGGAAGAAA-----ACTTGGGAAATGAAGCATGTTTGAATAATCA 1246  
DB 4435 TCTTGGCTCAAGTGAATTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4494  
QY 1247 GCAAGTGCAGGGTGTTCGGAGTTTATTTATATATATATATATATATATATATATATATAT 1306  
DB 4495 GCCACCATGCCAGCTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4552  
QY 1307 GATGAAGATCAAGTTATATCTAGTGAAGTGAATTAAGCTGGAATAGGCGTTCA 1366  
DB 4553 CATGCCCTGTAATCTTGAACCTTGAGAGCC-CGAGGACAGGAGGATTCCTGAGACTAGGA 4611  
QY 1367 GAGTAAATCATGAAGCACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAATA 1426  
DB 4612 GTTCAAAACCAACCTGCCACATAGCCAGTCTTTTAAATAATAATAATAATAATAATAATA 4671  
QY 1427 ATAAAAATCAAAATTTTTTTTTTTTTTTTGTAGAGAGGCTTGTCTTTTCCACCTGGCT 1486  
DB 4672 TTTATCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 4731

QY 1487 GGAGGCGAGTGTGTGATCTCAGCTCAGTCAACTTTGGCTCCCGGTTCAAGCAATTC 1546  
DB 4732 GAGTGCAGTGGGCGGATCTCAGCTCAGTCAAGCTCCGCTCTCTGGTTCACGCCATTC 4791  
QY 1547 TCTGCTTTCAGCTCCCAAGTAGCTGGGACTCAGGCACTTCCACCATGCCCCAGCTGA- 1605  
DB 4792 TCTGCTTTCAGCTCCGAGTAGCTGGGACTCAGGCTGCTGCCACCATGCTTGGCTAAT 4851  
QY 1606 --TTTTTGTATTTTGTAGTAGAGTGGGATTTCACTTTTGGCCAAAGCTGGTCTCAAACT 1663  
DB 4852 TTTTGTGTATTTTGTAGTAGAGGTTTCACTGTTTAAACAGGATGTTCTCAATCT 4911  
QY 1664 TTTTGTGTATTAATTTTGTAGTAGGTTTCTCTTTTGTGAGTGGGCCCCCAGACCA 1723  
DB 4912 CCTGAC-----TTCGTGATCCACCCACCTCAGCTCCCAAGTGTGGGATTCAGGCGT 4966  
QY 1724 AAAAAATAATTTCTTGTAGTAATCAATCACTGTTGTTGGTTGACCACTGTCTTGTAGAAC 1783  
DB 4967 GAGCCACCACTGCTGGCT-----GGCTATCTCTTTTAAACCTTTTATTTGGAGAA 5019  
QY 1784 CACAGTGTGACAGGCGCTCAGGAGTGTAGGTTGATCTCTGCTGCGAAGAGAAATAGATG 1843  
DB 5020 AAAAAATCAGAGTGCCTTTTGTGCTTTTACATGTCAGCAATTAAGTTGAAAAAATTTT 5079  
QY 1844 AAAATATTTCTCGGCGCAGGCGTGTGCTCATGCTGTAATCCCGACACTTTTGGAGGC 1903  
DB 5080 TTTTAAAGTGGGTGCTGGGCGGCTGCTCAGGCTGTAATCCCGACACTTTTGGAGGC 5139  
QY 1904 CAGGCACTGTGATCA-----CCTGAGTCTAGGAGT 1934  
DB 5140 TGAGCGCTGTGATCATGAGTTCAGGAGGCTGAGGAGTGCATCAACAAGTCCAGGAGA 5199  
QY 1935 TCAAAAACAGCTGCGCAACATGTTGAACCCCGTCTCTACTAAAAATACAAAAATTAG 1994  
DB 5200 TCGACACCATCTGCTAACGTTGTTGAACCCCATCTCTACTAAAAATAC----- 5258  
QY 1995 CTAAGTGTGTGGCGCATGCTGTAATCCAGTACTTCTGGAGGCTGAGGAGGAGATTT 2054  
DB 5259 CTGGCGTGTGTGCTGCTGTAATCCAGTACTTGTGAAGCTGAGGAGGAGGAGATTT 5318  
QY 2055 TCTTGAACCCGGGAGGAGGCTGCTGTAAGGAGATACACACCTGCTCAGCTT 2114  
DB 5319 GCTTGAACCCGGGAGGAGGCTGCTGTAAGGAGATACACACCTGCTCAGCTT 5378  
QY 2115 GGGGAGAGAGGAGGAGCTTCTCTCAAAAAACAAAAA 2153  
DB 5379 -GGCAACAGACAGACCTCTCTCAAAAAA 5416

## RESULT 2

US-09-797-906-3/c  
; Sequence 3, Application US/09797906  
; Patent No. 6329188  
; GENERAL INFORMATION:  
; APPLICANT: Zilanghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL001151CIP  
; CURRENT APPLICATION NUMBER: US/09797,906  
; CURRENT FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 84495  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)....(84495)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-797-906-3

Query Match 8.3%; Score 345; DB 4; Length 84495;  
Best Local Similarity 55.0%; Pred. No. 3.7e-74;  
Matches 979; Conservative 0; Mismatches 715; Indels 85; Gaps 12;

QY 445 AAACCTCCTGGGCTCAAGTATCCTACTACTAGCC-TCCAGAGTAGTAGGACTACAGG 503  
DB 18461 AACTCTGGGGTGAGCCATCCACCTGCTGGCCCTCCAAAGTGGGGATTACAAG 18402

QY 504 CACACACAGCCATACCTGGCTAAATTTTTTTTAAATTTTCATTTTATGATTCATTTTC 563  
DB 18401 CATAGTACACAGCCCTGGCATTTTATATATATATATATATATATATATATATTTT 18342

QY 564 TTTCTTTTGTG-TTGTTGTTGTTTTCAGATAGGCTCTACTTTGTTACCCAGGCTGGAG 622  
DB 18341 TTTTTTTTTTTTTTTTTTTTTTTGGACAGGGCTCTACTCTGTCCACCCAGGCTGGAGT 18282

QY 623 CGAGTGGCATGTGACAGCTGA--GCAGCCTTTGACTTCTCGCTGGCTCAAGTGATCCCTCCNG 680  
DB 18281 GCAGTGGCATGTCTCAGCTCACTGCAACTTGGACTCTGGCTCCAGCAATTCCTCCG 18222

QY 681 CCTCAGCCTCCCAAGTAGTGGGACTACAAACAGCTGTCACCATGCTGGCTGATATTTT 740  
DB 18221 CCTCAGCCTCTTAAGTAGTGGGATTTACAGGCATGTCACCGCGCCAGCTAAATTTTTT 18162

QY 741 TTTTCTTGAAACAGGGTATCACTCTGTGTCGCCAGGCTGGAGTACAGTGGCGTAAATATAG 800  
DB 18161 TTTTCTT-----TTTTTGTTTTTTTTTTTTGTATAGAGCGGAGTTTCACTATATATAG 18109

QY 801 CTCACGTGACGCTCCCTCTCTGGCTCAAGCAATCCGCTGGCTCAGCATCTCTGAGTAGC 860  
DB 18108 CCAGGCGCATCTCGAACTCTCAGCTCAAGTCACTGCTGCCGCTCAGCTCCCAAGTGT 18049

QY 861 TGGGACTACAGCTTGTGCCACAGCCAGCCAGCTAAGTTTTTAAAAATGATTTTTTGGTATA 920  
DB 18048 TAAGATTACAGCATGAGCCACCATCGCGGCTCTCTATTTTATAGTTTTTAAGACACAA 17989

QY 921 GAGGAGGTCTTCTATGTTGTCAGGCTGATTTTTTATGTTGACACAGGCTCTCACTAT 980  
DB 17988 GA----GTCCCACTCTGTGCCCAGGCTGGAGTGGAGTGGCCACATCATAGTCACTGCA 17933

QY 981 GTTGCCATGATGCCCCCACTCTCCCAAGTGCTCATCTTATCTG--TTCAATAGT 1038  
DB 17932 GCTTTAACTCTCGGCTAAAGCAATCCTCTACCTCAGCTCCCGAGTAGCTACATGAC 17873

QY 1039 CAGTTGACAGCATTTAGTGTGTTTCCATTTTGGACATATGATAATACTCCAGTGA 1098  
DB 17872 ATGCCACCATACCCAGCTAAATTTTTTAAATGTTTTTAAAGATGGGGTCTCACTATG 17813

QY 1099 ATATTCATGTATACATTTGTGTGGGCATATGTTTCATTTCTGTTGGTTTATATCTAGG 1158  
DB 17812 CCCAGGCTGGTCTGAACTCTCGGCTTAAGGATCCCTCCCATCTTGGCTTTCCAAAGTG 17753

QY 1159 AGTGAATGCTGGATCCCGGTAAATTTTACAGGCGAGGTTCCAGGGGAAGAAAACT 1218  
DB 17752 CTGGCATTTACAAGTGTGAGCCACCATGCTGGCCCAATGATCTTTTCAAGGTCACCT 17693

QY 1219 TGGGAATGAGCATGTTTGAATATCAGAGAGTGGAGGGTTTTTCGGAGTTTAT 1278  
DB 17692 GACTGCGATGG--AGAATAAATGGAGGAGTAAATAGTTTAAATAGGCGATTTTAA 17635

QY 1279 TTATATCTGTTGACAAATGTGCAAGTTTGAATAGATACAAGTTATATCTAGTGAAGT 1338  
DB 17634 ATGACCATACAACAGCCCATCTCACTTATGAAATAATCATTAATTCTGATCATTTCT 17575

QY 1339 GAGAAATTAAGGCTGGATFAGGCGCTCAGATGATAAATCATGACACTTTGA----ATAC 1394  
DB 17574 CTATTTTATGACATTTTATAGACTATTTCCAGTTCTTTCTTAATTAATAAAGCGAC 17515

QY 1395 CAAAATTAAGGCTGCTGTAAACAAATAATAAAAAATCAAAATTTTTTTTTTTTTT 1454  
DB 17514 TGTAAATCAATCTTGAACCTAAATCTTTGGCAGCATCTCTAATTAATTTCTTTCTTTT 17455

QY 1455 TTGAAAGAGCTTTGCTCTTTCCCTGCTGGTGGAGGCGAGTGGTGATCTCAGCTAC 1514

DB 17454 TTTTGACAGAGTCTCAGCTGTTGCCAGGCTCGAGTGCAGTGTGCAATCCCAAGCTCAC 17395

QY 1515 TGAACATTTGCGCTCCCGGTTCAAGCAATTCCTCTCAGCTCCCAAGTAGCTGGG 1574

DB 17394 TGAACCTCCACCTCCCTGGTTCAGTGAATCTCTCTCAGCTCCCAAGTAGCTGGG 17335

QY 1575 ACTACAGGCACTTCCACCATGCCAGCTGATTTTGTATTTTGTAGTAGAGTGGGATTT 1634

DB 17334 ATTACAGTGGGGCGCCACCATGCTGGCTAATTTTGTACTTTTAGCAGACGAGGTTT 17275

QY 1635 C-ACCTTGTGGCCAAAGCTGGTCTCAAACTTTT----- 1666

DB 17274 CGCAATTTTGGCCAGGCTGGTCTCAAACTCTGACCTCTGGGATTCACAGCGGTGAGCCA 17215

QY 1667 -----TGCTGTCAATTTGTTGTAATTTGTTTCCCTTTTGTGAGTGGGCCCCCA 1718

DB 17214 CCAGCGCCGACTTAATTTTCTAAATATAGATTTCTAGAGTGAATTAAGTAAATAA 17155

QY 1719 GACCAAAAAAATAAATCTTAGAATCCAAATCAGTGTGTTGTTTGCACCTGTCACCTTG 1778

DB 17-54 GTTAAAGGTATGAAATTTTGGCTATTTTATAAAATTTGTTTTTAAAGATAAAGATT 17095

QY 1779 AGAACACAGTGTGACAGGCGCTCAGAGTAGAGTGTATCTCTGCTCGAAAGAGAAATA 1838

DB 17094 TTGATCATTAATAATAAAAAATCTACTTTGCAAAATGGAATATCTTTTCACTATCCAGTC 17035

QY 1839 GAAT-----GAAATATTTCCGGGCCAGGCGGTGGTGGCTC 1874

DB 17034 AAATTATATCAGGGGAAATGTGCTTTTAAAGATGAAAACTAGGCCAGCGAGTGGCTG 16975

QY 1875 ATGCTGTATATCCAGCACTTTTGGAGGCGCAAGGCATGTGGATCACTGAGGTCAGGAGT 1934

DB 16974 AGCGCTGTATATCCAGCACTTAGGAGGCTGAGTGGGCGGATTAC--AAGTTCAGAGT 16917

QY 1935 TCAAAACAGGCTGGCCACATGTTGAACCCCGTCTCTACTAAAAATACAAAAATTAG 1994

DB 16916 TCGAGACCATCTGCTACATGTTGAACCCCATCTCTACTAAAACTAAAAAAATAG 16857

QY 1595 CTAAGTGTGGTGGCGCATGCTGTAATCCAGCTACTTTGGGAGGTTGAGGCGAGGAAAT 2054

DB 16856 CCGGTTGTCTGGGCGGCTGTAGTCCAGCTACTCGGAGGCTGAGGCGAGGAAAT 16797

QY 2055 TCTTGAACCCGGGAGCGAGGTTGCAGTGAAGCGAGATCACACCACTCACTCCAGCT 2114

DB 16796 TCTTGAACCCGGGAGATGAGGTTGCAGTGAAGCGAGATTGACCACTCACTCCAGCT 16737

QY 2115 GGGGAGAGAGGAGACTTCTCTCAAAAAACAAAAA 2153

DB 16736 GGGCAACAGGCGCAGACTCTGATTCAAAAAAGAAAAA 16698

## RESULT 3

US-08-814-095-7  
; Sequence 7, Application US/08814095  
; Patent No. 6025183  
; GENERAL INFORMATION:  
; APPLICANT: Sored, Hermona  
; APPLICANT: Zakut, Haim  
; APPLICANT: Shani, Moshe  
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KOHN & ASSOCIATES  
; STREET: 30500 No. 6025183thwestern Highway, Suite 410  
; CITY: Farmington Hills  
; STATE: Michigan  
; COUNTRY: U.S.  
; ZIP: 48334  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

```

?
?
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: 2391.00066
? FILING DATE: US/08/814,095
? CLASSIFICATION: 800
? ATTORNEY/AGENT INFORMATION:
? NAME: Montgomery, Ilene N.
? REGISTRATION NUMBER: 38,972
? REFERENCE/DOCKET NUMBER: 2391.00066
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (248) 539-5050
? TELEFAX: (248) 539-5055
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 35060 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "Cosmid including ACHE
? DESCRIPTION: promoter, ACHE gene and ARS gene"
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Homo sapiens
? POSITION IN GENOME:
? CHROMOSOME/SEGMENT: 7q22
? FEATURE:
? NAME/KEY: exon
? LOCATION: 22465..22537
? OTHER INFORMATION: /function= "ACHE Promotor"
? OTHER INFORMATION: /standard_name= "ACHE Promotor"
? FEATURE:
? NAME/KEY: exon
? LOCATION: 24090..25177
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /function= "(translation start:
? OTHER INFORMATION: 24110)"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 2
? FEATURE:
? NAME/KEY: exon
? LOCATION: 25524..26009
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 3
? FEATURE:
? NAME/KEY: exon
? LOCATION: 27005..27274
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 4
? FEATURE:
? NAME/KEY: exon
? LOCATION: 27255..28007
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 5
? FEATURE:
? NAME/KEY: terminator
? LOCATION: 27385..27387
? FEATURE:
?
?
? NAME/KEY: exon
? LOCATION: 28008..28129
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 6
? FEATURE:
? NAME/KEY: terminator
? LOCATION: 28129..28131
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (34528..34895)
? OTHER INFORMATION: /function= "arsenite resistance
? OTHER INFORMATION: gene"
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 1
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (34092..34358)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 2
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (33779..33963)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 3
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (33493..33591)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 4
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (33297..33408)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 5
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (32959..33094)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 6
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (32569..32628)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 7
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (32386..32468)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 8
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (31894..32080)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 9
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (31363..31534)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 10
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (31131..31284)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 11
? FEATURE:
? NAME/KEY: exon
? LOCATION: complement (30816..31011)
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 12
? FEATURE:
? NAME/KEY: exon

```

```

, , LOCATION: complement (30470..30626)
, , OTHER INFORMATION: /gene="AR"
, , OTHER INFORMATION: /number= 13
FEATURE:
, , NAME/KEY: exon
, , LOCATION: complement (30187..30274)
, , OTHER INFORMATION: /gene="AR"
, , OTHER INFORMATION: /number= 14
FEATURE:
, , NAME/KEY: exon
, , LOCATION: complement (29945..30073)
, , OTHER INFORMATION: /gene="AR"
, , OTHER INFORMATION: /number= 15
FEATURE:
, , NAME/KEY: exon
, , LOCATION: complement (29664..29856)
, , OTHER INFORMATION: /gene="AR"
, , OTHER INFORMATION: /number= 16
US-08-814-095-7

```

Query Match 7.7%; Score 321; DB 3; Length 35060;  
Best Local Similarity 68.8%; Pred. No. 1.8e-68;  
Matches 511; Conservative 0; Mismatches 205; Indels 27; Gaps 4;

QY	1445	TTTTTTTTTTTGAAAGAGTCTTGCTCTCTTCAACCTGGCTGGAGGAGTGGTGTGAT	1504
Db	9802	TCTTTTTCTTTTGAGCAGGAGTCTCGCTCTATCGCCAGGCTGAGTGGCACAAT	9861
QY	1505	CTCAGCTCAGTGCACACTTTCGCTCCCGGTTTCAAGCAATTTCTCCGCTTCAGCTCCCA	1564
Db	9862	CTCGGCTCAGTGCACACTTCTGCCACCCAGGTTCAAGCAATTTCTCTGTGTCAAGCTCCCA	9921
QY	1565	AGTAGCTGGAGCTCAGAGCACTTCCCAACCAATGCCAGCTGATTTTGTATTTTATTAGTAGA	1624
Db	9922	ANTAGCTGGGATTACCAAGTCCACACACATGCCAGCTAATTTTATATTTTTCAGTAGA	9981
QY	1625	GATGGGATTTCACTTTGTGTGGCCAAAGCTGTGTCTCAAACTTTTGTGTCTATAATTGTGT	1684
Db	9982	GACAGGGTTTCACCATGTGTGCCAGGCTGTCTCAAAATCTCGCTCAAGTGATCTGCC	10041
QY	1685	AAC-----TATTGTCTCTTTTGCTGAGGTAGGCCCCCAGACCCAAAAAATA	1732
Db	10042	CACCTCAGCCTCCCAAGTGTGTGGATTACAGATGTGAGCCACCTCACTCAGCCAGAAATC	10101
QY	1733	AATCTTAGAATCCCAATCAGTGTGTGGTTTGACCACTGTCACTTCAGAAACCA	1785
Db	10102	CTTCTCCAATTTAAATCAACAGAAACCCACAGAAATCACATTTGTTGGCATTGCTCT	10161
QY	1786	-----CAGTGTGACCAGGGCCTCAGGAGTAGAGGTGATCTCTGCTGAAAGAGAAATAG	1839
Db	10162	AGTAATATGTGCATCTTGGGTGCAAGGCTCTAAATGCTAAACCTGCATCTCAGTTTTA	10221
QY	1840	AATGAAATATTTCTCCGGCCAGCGGTGGCTCATGCTTAATCCCAAGCACTTTGGG	1899
Db	10222	CATTTTAAGTGTCTCCGGCCAGCGGTGGCTCAGGCTGTAAATCCCAAGCACTTTGGG	10281
QY	1900	AGGCCAAGGCATGTGGATCACTTGAGTCAAGGAGTTCAAAACCAAGCTGGCCCAACATGCT	1959
Db	10282	ANCGCAGGCAGGTGGATCACTTCAGTCAAGAGTTCGAGATAGCCTGACCCACATAGT	10341
QY	1960	GAACCCCGTCTCTACTAAAAATACAAAAATTAGCTAAGTGTGGCGCATGCTGTA	2019
Db	10342	GAACCCCATCTACTAAAAA-ATACAAAAATTAGCCAGGTGTGGTGGCGATGCTGTA	10400
QY	2020	ATCCCACTACTTGGAGGGTGGAGCAGGAGAAATTTCTTGAACCCGGGAGGCAGAGTTG	2079
Db	10401	ATCCCACTACTTCAGGAGGCTTGAGGCAGAGAAATTCGCTTGAACCCGGGAGGCAGAGATTG	10460
QY	2080	CAGTGAAGCGAGATCACACACTTCGACTCCAGCTCGGGGAGA-GAGCGAGACTTCCTCT	2138
Db	10461	TAGTGAGCTGAGATGCACCATGCGACTCCTCCTGGGCAACAGGAGCAAAATCCGCTCT	10520
QY	2139	CAAAAAACAAAAAACAAAAAGAA	2161

```

||||| ||||| ||||| ||||| |||||
Db 10521 CAAACAAACAAACAAACAGAAA 10543

RESULT 4
US-09-813-817-3/c
; Sequence 3, Application US/09813817
; Patent No. 6340583
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN K
; TITLE OF INVENTION: ACID MOLECULES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: GL001178
; CURRENT APPLICATION NUMBER: US/09/81
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
US-09-813-817-3

```

Query Match	7.6%	Score 316.6;	DB 4;	Length 59065;
Best Local Similarity	66.7%;	Pred. No. 2.8e-67;		
Matches 561;	Conservative	0;	Mismatches 264;	Indels 16;
			Gaps	7;

QY	1339	GAGAAATTAAGGCTGGAAATAGGGCGTTCAGAGTAAATCATGAAGACACTTTGTAATACCAA	1398
Db	12464	GAGAGGTGAGGCTGAAGTGCAGGAGCAGGAGGTTGTGGAGATGTGCAGAGAGCCACAT	12405
QY	1399	ATTAAAGGAGCTGGCGTGTAAACAAATAAATAAAATACAAATTTTTTTTTTTTTTTTGA	1458
Db	12404	TACGGAGGCTCTACTGTAGGCCAAATTA-AGGAATTTTTTTTTTTTTTTTTTTTTTGA	123461
QY	1459	GAAGAGCTTGTGCTCTTTCAACCTGGGTGGAGGACAGTGGTGTGATCTCAGCTCACTGCA	1518
Db	12345	GATGGAGCTTACTCTGTGCACA-GGCTGGAGTGCAGTGGCGTGATCTGGGCTCACTGCA	12287
QY	1519	ACTTTTCGCTCCCGGTTCAAGCAATTCCTGCTTCAGCCCTCCCAAGTAGCTGGGACTA	1578
Db	12286	ACCTTGCTCCCGAGTTCAAGTGATCTCTGCTCAGGCTCCCAAGTAGCTGGGACTA	12227
QY	1579	CAGGCACCTCCACACATGCCAGCTGATTTTTGTATTTTAGTAGAGATGGGATTTCACT	1638
Db	12226	CAGGCACACGCCACCACGACGACTAATTTTGTATTTTAGTAGAGATGGGTTTCACC	12167
QY	1639	TTCGTTGGCAAGCTGGTCTCAAACTTTTTGCTGTGCATAAT-----GTTGTAACT	1688
Db	12166	ATGTTGGCCAGATGGTCTCAATCTCTTGACTTCATGATGCTGCCCTCGGCCCTCCCAA	12107
QY	1689	ATTGTTCCCTTTTGTGTAGGTAGGCGCCCGCCAGACCAAAAAATAAATCTTAGAATCCAAA	1748
Db	12106	AGTGTCTGAGATGATAGGTGTGAGCCACCGCGCTTGCCCAAGTAAGGATTTTGGTCTTA	12047
QY	1749	TCAGTGTGTGGTTTGACCACTGTCACTTGAGAACCACTGTGTGACAGGGGCTCAGGAG	1808
Db	12046	TCCAAGAACACTGGGAAGTTGTGGAAGTATACAAAGTAGGAGAGACACACAGAGTGACA	11987
QY	1809	TAGAGGTGATCTCTGCTCGAAGAGAAATAGATGAATATTTCTCCGGGCCAGCGCTGG	1868
Db	11986	CAGTCAAGTTTTTTTTTGTTTTTTTTTTTTTTAAATCA-CTCCTTCAGCCAGGTGCAG	11928
QY	1869	TGGCTCATGCTGTAAATCCAGCACTTTGGAGGCCAAGGCATGTGGATCACTGAGTTC	1928
Db	11927	TGGCTCACACCTGTAAATCCAGCACTTTGGAGGCCGAGCAGGTGGATCACTTGAGTTC	11868
QY	1929	AGGAGTTCAAAACACAGCGTGCACCAATGGTGAACCCCGTCTCTACTAAAAATACAAA	1988
Db	11867	AGGAGTTCGAGACTAGCTTGGCCAAATGCAAAACCCCATCTCTACTAAAAATA-TAAA	11809
QY	1989	AATTAGCTAAGTGTGGTGGCGCATGCGCTGAATCCACGACTCTTGGGAGGGTGGAGCAGG	2048



Db	11808	AATTAGCTGGGTATGGGGGCGGTGCCT-TAATCCAGCTACTTTGGGAGGCTGAGCAGG	11750
QY	2049	AGAAATTTCTTGAACCCGGGAGGAGAGAGTTGCAGTGAAGCGAGATCACACCACCTGCACCTC	2108
Db	11749	AGAAATGTCTTAAACCCGGGAGGAGAGAGTTGCAGTGAGCCGAATTTGTCCACTGCACCTC	11690
QY	2109	CAGCC-TGGGGGAGAGCGAGACTTCTCTCAAAAAACAACAAAAACAAGAATTTAGC	2167
Db	11689	CAGCCTTGGGGAACAGAGTGAGACTCTGTCTAAAAACAACAAAAACAACAAAAATC	11630
QY	2168	A 2168	
Db	11629	A 11629	

```

RESULT 5
US-09-978-197-3/c
; Sequence 3, Application US/09978197
; Patent No. 6403353
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEOTIC
; MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, NUCLEOTIC
; MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001178DIV
; CURRENT APPLICATION NUMBER: US/09/978,197
; CURRENT FILING DATE: 2001-10-17
; PRIOR APPLICATION NUMBER: 09/813,817
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
; US-09-978-197-3

```

Query Match	7.6%	Score 316.6	DB 4	Length 59065
Best Local Similarity	66.7%	pred. No. 2.8e-67		
Matches 561	Conservative 0	Mismatches 264	Indels 16	Gaps 7
QY 1339	GAGAAATTAAAGCTGGATAGGGCGCTCAGAGTAAATCATGAAGACACTTTGAAATACCAA	1398		
Db 12464	GAGAGGTGAGGCTGAAGTGCAGGGAGCAGGAGGTTGIGGAGATGGAGAGAGGCCACAT	12405		
QY 1399	ATTAAAGAGCTTGCTGTAAACAAATAATAAARAAATCAACAATTTTTTTTTTTTTTTGA	1458		
Db 12404	TACGGAGGGCTCTACTGTAGGCCAAATTA-AGSAAATTTTTTTTTTTTTTTTTTTGA	12346		
QY 1459	GAAGAAGTCTTGCTCTTTACCCCTGGGTGGAGGCGAGTGGTGATCTCAGCTCACTGCA	1518		
Db 12345	GATGGAGTCTTACTCTGTACCA-GGCTGGAGTGCAGTGGGCGTCACTGGGCTCACTGCA	12287		
QY 1519	ACTTTGGCTCCCGGTTCAAGCAATTCCTCGTTCAGCCTCCCAAAGTAGCTGGGACTA	1578		
Db 12286	ACCCTCGCTCCCAAGTTCAAGTAGTATCTCGCTCAGCCTCCCAAGTAGCTGGGACTA	12227		
QY 1579	CAGCACTTCCCAACATGCCAGCTGATTTTGATTTTATAGTAGAGATGGATTTCACT	1638		
Db 12226	CAGCACACGCGACCCAGCAGCCAGCTAATTTTGTATTTTATAGTAGATGGGTTTCA	12167		
QY 1639	TTGTTGGCAAGCTGGTCTCAAACTTTTGGTGTCAATAAT-----GTTGTA	1688		
Db 12166	ATGTTGGCAGGATGGTCTCAATCTCTGTGCTCATGATCTGCCTGCCTGCCCTCCAA	12107		
QY 1689	ATTGTTCTTTTGTGAGTAGGGCCCCCAGACCAAAAAATAAATCTTAGAATCCAAA	1748		
Db 12106	AGTGCTGAGATGATAGTGTGAGCCACCGCGCTGGCCAAAGTTAAGGATTTTGGTCTTTA	12047		
QY 1749	TCAGTGTGTTGTTGACCACTGTCACTTGAGAACCAACACAGTGTGACCAAGGCCCTCAGGAG	1808		
Db 12046	TCCAAGAACAACCTGGGAAGTTGTGGAAGCTATACAAGGTAGGAGAGACAAACACAGAGTGACA	11987		

```
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
```

Query Match	7.48;	Score 308.2;	DB 4;	Length 162450;
Best Local Similarity	66.7%;	Pred. No. 5.3e-65;		
Matches 529;	Conservative 0;	Mismatches 233;	Indels 31;	Gaps 5;
QY 1421	AAAAATAAATAATACAAATTTTTTTTTTTTTTTTTTTGAGAGAGAGCTGTGCTCTTTTACC	1480		
Db 86851	AATCTAAACAAGTGAACAACTATTTTTTTTCTTTTGAGCGGAGCTAGCTCTGT-TGC	86793		
QY 1481	CTGGCTGGAGGGCAGTGGTGTAATCTCAGCTCACTGCAACATTTGCCCTCCGGGTTCAAG	1540		
Db 86792	CGGGCTGGAGTGCAGTGGCACACACTTGTGCTCTCACTGCAAGCTCGGCCCTCTGGGTTCAAG	86733		
QY 1541	CAATTCCTCTGCTTTCAGCTCCCAAGTAGCTGGGACTACAGGCACCTTCCCAACCATGCCCA	1600		
Db 86732	CAATTCCTCGCCTCAGCCTCCTGAGTAGCTGGGATACAGGTGCTCGCCACCGCCCA	86673		
QY 1601	GCTGATTTTGTATTTTAGTAGAGATGGGATATTCACPTTGTGGCCAAAGTGTCACAA	1660		
Db 86672	GCTAAITTTTGTATTTTGTAGAGAGGGGTTTCCACCATGTAGCCAGAGATAGTCTCGA	86613		
QY 1661	ACT-----TTTGTCTGTCAATAATGTTGTTACTACTATGTTCTCTTTGCTGAGGT	1708		
Db 86612	TCTCTGACCTCGTGATCCGGCCCACTCGCCTCCCAAGTCTGGGATATAGCGGTGAG	86553		
QY 1709	AGGGCCCCAGACCAAAAAAATAAATCTTAGAATCCAAATCAGTGTGTTGGTTTGACCA	1768		
Db 86552	CCACCCACCTGGCCAAAAAAGTAAAACTATTATATAATATTAAAGCTTTTAAAGTGA	86493		
QY 1769	CTGTCACTTGAGAACACACAGTG-----TGACAGGGCCCTCAGAGTAGAGTGATCTC	1821		
Db 86492	CATGGGAAATACCCACCCCTTGATTTCAATGATCAGTCAAAATACTTAAAAATTTGTC	86433		
QY 1822	TGCTCGAAAGAGAAATAGAAATGAAATATCTCGGGCCAGG-----CGTGGTGG	1871		
Db 86432	TGGCAGAAACAACAATCTTCAAGCTTGCTATTGATAGAGTAGTGTACTTTGGCT	86373		
QY 1872	CTCANGCCTGTATCCAGCACATTTGGAGGCCAAGGCATGTGATCACCTGAGGTCAAG	1931		
Db 86372	GGAAGTATTGTACTCCAGCACATTTGGAGGCGAAGCAGGTGATCACCTGAGGTCAAG	86313		
QY 1932	AGTTCAAACACAGCTCGGCCAACATGTGAAACCCCGTCTCTACTTAAAAATACAAAAAT	1991		
Db 86312	AGTTCGAGACAGCCTGGCCAAAATGTTGAACTCTGCTCTCTCTTAAAAATAC-AAAA	86254		
QY 1992	TAGCTAAGTGTGGTGGCGCATCGCTGTAAATCCCAAGCTACTTTGGAGGGTGAGGCAGGAG	2051		
Db 86253	TAGCCAGGCATGTGGCGGGTGCTGTAAATCCCAAGCTACTCTGGAGGCTGAGGCAGGAG	86194		
QY 2052	ATTTCCTTGAACCCGGAGGAGAGGTTGCAGTGAAGCGGAGATCACCACCTGCCTCAG	2111		
Db 86193	ATTGCCCTGAACCCGGAGGCGGAGGTTGCAGTGAAGCGGATATCCACACCTGCCTCAG	86134		
QY 2112	CCTGGGGAGAGCGAGACTTCCTCTCAAAAAACAAAAACAAAGAAATTAAGCAAT	2171		
Db 86133	CCTGGCGACAGTAGACTCCGCTCAAAAAAATAAAAAAAGATGTATTGTACTTCAA	86074		
QY 2172	TAGACATTTGCAGA 2184			
Db 86073	TGGCACTTAAAGA 86061			

		Query Match	7.4%; Score 307.6; DB 4; Length 43950;	
		Best Local Similarity	67.6%; Pred. No. 3.9e-65;	
		Matches 491; Conservative	0; Mismatches 229; Indels	6; Gaps 4;
QY	1442	TTTTTTTTTTTTTTGAGAAAGACTCTGTCTTTTACCGTGCGTGAGGGCAGTCGTGT	1501	
Db	9962	TTTTTTGTGTGTGAGACAGACTCTGTCCGTACCCAGATGAGTGCAGTGGCGT	10021	
QY	1502	GATCTCAGCTCACTGCACAATTTCGCTCCGGSTTCAAGCAATTTCTCTGTTTCAGCCTC	1561	
Db	10022	GGTCTTGGCTCACATGCAACCTCTACCTCTGGGTTCAAGCAATTTCTCTGTTTCAGCCTC	10081	
QY	1562	CCAAGTAGCTGGGACTACAGCAGCTTCCCACCATTGCCAGTGCAGTGTATTTTGTATTTTAGT	1621	
Db	10092	CCAAATAGCTGGGATTACAGGTGGCGCACCACTCCTGGCTAATTTTGTATTTTGTATTTTAGT	10141	
QY	1622	AGAGATGGGAATTCACATTTGTGTGCCAAGCTGGTCTCAAACCTTTTTCGTGTCATAAATGT	1681	
Db	10142	--AGACGGGGTTTCACCATGTTGGCCAGGTGGTGTTCAGTAATCCACCTCTCCFCAGCCT	10199	
QY	1632	TGTAACATATTGTTCTCTTGTGTAGGTAGGGCCCCCAGACCAAAAAAATAAATCTTAGA	1741	
Db	10200	CCCCAAGTGTGGGATTACAGACATGAGCCACTGCGCTGGGCCCATCGCGCTCCATTTT	10259	
QY	1742	ATCCAAATCAGTGTGTGGTTTGACACTGTCACTGTAGAACACACAGTGTGACACAGGGCC	1801	
Db	10250	TGTATGTTGTGCCCTGCAATTAGAGCCATATTCCTTGCATGTTCCATTGGGTATTAGTCT	10319	
QY	1802	TCAGGAGTAGAGTGTATCTCTGCTCGAAGAGAAATAGATGAANAATTTCTCCGGGCCA	1861	
Db	10320	GAGACAGCATCTCTAGCTCCGTGGGTGCCACGCTTGTACAGAAATCTCGATTCTGGGCCA	10379	
QY	1862	GGCTGTGTGGCTCATGCTGTATAATCCAGCAGCTTTGGGAGGCCAAGGCATGTGGATCAC	1921	
Db	10380	GGCAGGTGGCTCACACTGTATCCAGCAGCTTTGGGAGGCCAAGGGGGGGGATCA--	10437	
QY	1922	TGAGGTCAGGAGTTTCAAAACCCAGCTGGCCACACATGGTGAACCCCGTCTCTACTAAAA	1981	
Db	10438	TGAGGTCAGGAGTTTAGAGACCAAGCTTGGCCAAACATGGTGAACCCCTGCTCTACTAAAA	10497	
QY	1982	TACAAAAAATTAGCTAGTGTGGTGGCGCATGCTGTAAATCCAGACTACTTGGAGGGTGC	2041	
Db	10498	TAGAAA-AAATAGCTGGGTGTGGTGGCGGTACTCTAATATCCAGCTACTCTGGGAGGCTG	10556	
QY	2042	AGGACGAGGAATTTCTTGAACCCG-GGAGGCAGAGTTGCAGTGAAGCAGAGATCACACCA	2100	
Db	10557	AGGACGAGGAATCATTTGACCTTGAGGGGTGGAGGTGCAGTGCAGCCGAGATCATACCA	10616	
QY	2101	CTGCACCTCCAGCCTGGGGGAGAGCGGAGACTTCTCTCAAAAAAACAAAAACAAAGA	2160	
Db	10617	TTGCACCTCCAGCCTGGGTGACAGGTTGAGACTCCGTCTCAAAAAAATAAAAAAAGAA	10676	

QY 2161 ATTAAG 2166  
|||  
Db 10677 ATCCAG 10682

## RESULT 8

```

RESULT 9
US-09-735-934A-3/C
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:
; APPLICANT: LI, JIAYIN et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000851
; CURRENT APPLICATION NUMBER: US/09/735,934A
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-735-934A-3

```

Query Match	7.3%	Score 303;	DB 4;	Length 43950;
Best Local Similarity	68.2%;	Pred. No. 5.3e-64;		
Matches 492; Conservative	0;	Mismatches 200;	Indels 29;	Gaps 4;

1441	QY	TTTTTTTTTTTTTTTTTTGAGAAAGAGTCTTCTCTTTTACCCCTGSCCTGAGGCGCAGTGGT	1500
6167	Db	TTTTTTCTTTCTGCTGTTGTTAGAGAGTTTCACTGTGTACCCAGCGCTGAGTACAGTAGTG	6108
1501	QY	TGATCTCAGCTCACTGCAACTTTCGCTCCCGGTTTCAAGCAATCTCCCTGCTTCAGCCT	1560
6107	Db	AGATCTGGCTCACCAAA-GTCCGCTCTCCGGTTCAGCAGTTCTCTCTGCTCAGCCT	6049
1561	QY	CCCAAGTAGCTGGAGCTACAGGCACCTTCCACCATGCCAGCTGATTTTGTATTTTAG	1620
6048	Db	CCAGAGTAGCTGGGATATACAGGCATGGCCACCACACCCGCACTAATTTTGTATTTTAG	5989
1621	QY	TAGAGATGGGATTTCACTTTTGGCCA-AGCTGGCTCAAACTTTTGTCTGTCATAATT	1679
5988	Db	TAGACGAGAGTTTCACTATTTGGCCAGGATGGTCTCAAACTCCGAGCTCAGGTGATC	5929
1680	QY	GTGTAACTATGTTCTTTTGTGTAGGTAGGGCCCCAGACCAAAAAAATAAATCTTA	1739
5928	Db	TGCCCGCTCGGTCTCTCAAGTGTGGGATTACAGCGGTGAGCCACTGCACCCGGCCTA	5869
1740	QY	GAATCAAATCAGTGTGTGGTTTGACCACTGTCACTTGAGAACCCACAGGTGACCAAGG	1799
5868	Db	AAATAGTAATTTTATGTTATGTATATTTTACCAATTAATAATTTTAAATTAAATA	5809
1800	QY	CCCTCAGGAGTAGAGTGATCTCTGCTGAAGAGAAATAGATGAAAAATATTCCTCGGGC	1859
5808	Db	CC-----ACCACCACCAAAAAAGAACATGATGATAGGC	5775
1860	QY	CAGGCGTGTGGCTCATGCTGTAATCCAGCACTTTGGGAGGCCAGGCATCTGGATCA	1919
5774	Db	GGTCCCGGTGGGCTCAGTCTGTAATCCAGCACTTTGGGAGGCCAGGTGGGCGGATCA	5715
1920	QY	CCCTGAGTTCAGGAGTTCAAAACAGAGCTGGCCAACTGTGTGAACCCCGCTCTACTATA	1979
5714	Db	CTTTGAGTTCAGAAAGTTCGAAACAGAGCTGGCCAACTGATGAACCCCGTCTACTATA	5655
1980	QY	AATACAAAAATTTAGCTAAGTGTGGGCCCATGCTCTGTAATCCAGCTACTTTGGGAGG	2039
5654	Db	AATAC-AAAAATTTAGCCAGGCTTGTGGCGGTGCTGTAATCCAGCTACTCAGGAGGC	5596
2040	QY	TGAGGCGAGGAGAAATTTCTGAACCCGGGAGGCAGAGTGTGCATGAAGCGGATCATACC	2099

Qy 2161 ATTAAG 2166  
 == ==  
 Db 10677 ATCCAG 10682

## RESULT 8

```

US-09-851-896-3/C
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEPENDENT)
; TITLE OF INVENTION: EXPRESSION
; FILE REFERENCE: R1S-0220
; CURRENT APPLICATION NUMBER: US/09/851,896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-851-896-3

```

Query Match 7.4%; Score 307.2; DB 4; Length 70000;  
Best Local Similarity 70.2%; Pred. No. 6.2e-65;  
Matches 504; Conservative 0; Mismatches 193; Indels 21; Gaps 6;

QY	1441	TTTTTTTTTTTTTTTGAGAAAGAGTCTTGGCTTTTACCCCTGGCTGGAGGCGAGTGGTG	1500
Db	5040	TTGTTTTGTTTTTTTTTGAGACGGAGTCTCATCTGTGCACCCAGGCTGGAGTGCATGGCG	50341
QY	1501	TGATCTCAGCTCACTGCAAACTTTGCGCTCCCGGGTTTCAAGCAATTCCTCGTTCAGCCT	1560
Db	50340	TGATCTCAGCTCACTGCAAACTCTACCTACCGAGTTCAGCAATTCCTCGCTCAGCCT	50281
QY	1561	CCCAAGTAGCTGGACTACAGGCACCTCCACCAATGCCACAGTGA-TTTTTGTATTTTTTA	1619
Db	50280	CCAGAGTAGCTGGTATTACAGTGGCCACCACTACGCCACGCTAAATTTTTGTATTTTTTA	50221
QY	1620	GTAGAGATGGGATTTCACCTTTGTGGCCAAGCTGGTCTCAAACTTTTGTGTGTCATATT	1679
Db	50220	GTAGACACAGGGTTCCACCAATGTTGCCAGGCTGGTCTGAAC---TCTGACCTCGTG	50165
QY	1680	GTGTGAACATATTGTCCTTTTGTGAGGTAGGGCCCCGACGACCAAAAAATAATCTTA	1739
Db	50164	ATTGGCCCACTCGGCCCTCCAAAATGCTGGGATTACAGCGCTGAGCCACGCGCCCT--	50107
QY	1740	GAATCCAAATACAGTGTGTGGTTTGACCACTGTCACTGTGAGAACCAAGTGTGACCAAGG	1799
Db	50106	-----GCATTGGTTACCTTTTATGGTTGGCACT--GGATGGATGGATGGCCATTTA	50058
QY	1800	CCTCAGGAGTAGAGTGATCTTGCTCGTGAAGAGAAATAGAAATAATTCFCCGGCG	1859
Db	50057	TGCTAGTGATGCAACATTTCTCTCAAAATAAATTTAAGTCAGTTTAAAGTAAGCCAGCG	49998
QY	1860	CAGCGCTGFGGGCTCATGCGCTGTAATCCCAAGCACTTTGGGAGGCCAAGGATGTGGATCA	1919
Db	49997	CGGCGACGGTGGCTCAGCTTTGTAATCCCAAGCACTTTGGGAGGCTGAGCGAGTGGAAACA	49938
QY	1920	CCTGAGTCAAGGAGTTCAAAACAGCGCTGGCCCAATAGTGTGAACCCCGCTCTCTACTAAA	1979
Db	49937	CCTGAGGTCAGGAGTTTCAAGACCAAGCGCTGACCAACATGAGAAACCCCGCTCTCTACTAAA	49878
QY	1980	AATACAAAAAATTAGCTTAAGTGTGGTGGCGCATGCCCTGTAAATCCACGCTACTCTGGGAGG	2039
Db	49877	AATAC--AAATTTAGCCAGTGTGGTGGCGCATGAACGTAATCCACGCTACTCGGGAAC	49820
QY	2040	TGAGCGAGGAGAAATTTCTTGAACCCCGGAGGCGAGGTTCCAGTGAAGCGAGCATCACCC	2099



```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
;
US-08-724-394A-21

Query Match          7.3%; Score 301.8; DB 2; Length 246240;
Best Local Similarity 68.0%; Pred. No. 2.4e-63;
Matches 493; Conservative 0; Mismatches 177; Indels 55; Gaps 3;

QY 1441 TTTT...TTTGTGAGAAAGAGTCTTGTCTTTTCACCCCTGGCTGGAGGCGAGTGGTG 1500
      |||...|||
DB 2626 TTA...TTTATTTTGTAGAGCTTGTCTCACCTTCGCCCGAGGCTGAGTGCAGTGGTG 2685

QY 1501 TGA...TCTAGCTCAGTCAACTTTTCCCGGGTTCGAAGCAATTCCTCGCTTCAGCCT 1560
      |||...|||
DB 2686 CAG...TCTGGCTCACTGCAATCCCGAGTCCCGAGGTTCAAGGATTCCTCGCTCAGTCT 2745

QY 1561 CCC...AGTAGTGGGACTACAGGCACTTCCCGAGGCTGGCTGCTCAAACTTTTCTGTCTATAATTG 1620
      |||...|||
DB 2746 CCC...AGTAGTGGGACTACAGGCTGGCTGCTCAAACTTTTCTGTCTATAATTG 2805

QY 1621 TAG...ATGGGATTTACCTTTGTGGCCAGCTGCTTCAAACTTTTCTGTCTATAATTG 1680
      |||...|||
DB 2806 TAG...ACAGAGGTTTGGCGGCTGTGGCCAGGCTGGTCTGCACT----- 2848

QY 1681 TGT...AACTATTGTCCTTTTGTGAGGTAGGGCCCCCGAGCAAAATAAATCTTTAG 1740
      |||...|||
DB 2849 -----CCTGAGCTCAAGTGATCTACCCGCC 2873

QY 1741 AAT...CAATCAGTGTGTTGTTGACCACTGTCTTGCAGAACCACTGTGACCAAGGC 1800
      |||...|||
DB 2874 TCG...GCTCCCAAGTGTGGGATACAGTGTGAGCCAGTGGCCCGCCCGGGGTTTATC 2933

QY 1801 CTC...AGGAGTAGAGTGATCTCTGCTCGAAGAGAAATAGAAATATTCTCCGGGCC 1860
      |||...|||
DB 2934 CTT...AATA-CATACATTAGAGATATAGATCTCTTTTATCTATAAAGCTTTATAAGGCC 2992

QY 1861 AGG...CGTGTGCTATGCTGTATCCAGCACTTTGGGAGGCCAAGCGATGTGGATCAC 1920
      |||...|||
DB 2993 GGG...CGCGTGTGCTACAGCTCTTAATCCAGCACTTTTGGAGGCCCGAGCGGGGATCAC 3052

QY 1921 CTG...AGTCTAGAGTTCAAAACAGCGCTGCGCAACATGTGAAACCCCTCTCTACTAAA 1980
      |||...|||
DB 3053 --G...AGTCTAGAGATCTGAGACCATCTCCCGCTTAAACCGTGAACCCCGCTCTACTAAA 3110

QY 1981 ATA...AAAAATTAGTAAGTGTGGTGGCGCATGCTGTAAATCCAGCTACTTTGGAGGGT 2040
      |||...|||
DB 3111 ATA...AAAAATTAGCGGCGTAGTGGCGGCGCTGTAGTCCAGCTACTTTGGAGGCT 3170

QY 2041 GAG...GAGGAGAAATTTCTTAAACCGGGAGGAGAGGTTGCAAGTGAAGCGAGATCACACCA 2100
      |||...|||
DB 3171 GAG...GAGGAGAAATTTCTTAAACCGGGAGGAGAGGTTGCAAGTGAAGCGAGATCACACCA 2160

QY 2101 CTG...CACTCCAGCTGGGGAGAGAGCGAGACTTCTCTCAAAAAACAAAAACAAAAAGA 2160
      |||...|||
```

```
DB 3231 CTG...CACTCCAGCTGGGCGACAGACAGACTCCGCTCTCAAAAAACAAAAACAAAAA 3290
QY 2161 ATT...AA 2165
DB 3291 AAA...AA 3295

RESULT 12
US-08-724-394A-22
; Sequence 22: Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
;
US-08-724-394A-22

Query Match          7.3%; Score 301.8; DB 2; Length 246240;
Best Local Similarity 68.0%; Pred. No. 2.4e-63;
Matches 493; Conservative 0; Mismatches 177; Indels 55; Gaps 3;

QY 1441 TTTT...TTTGTGAGAAAGAGTCTTGTCTTTTCACCCCTGGCTGGAGGCGAGTGGTG 1500
      |||...|||
DB 2626 TTA...TTTATTTTGTAGAGCTTGTCTCACCTTCGCCCGAGGCTGAGTGCAGTGGTG 2685

QY 1501 TGA...TCTAGCTCAGTCAACTTTTCCCGGGTTCGAAGCAATTCCTCGCTTCAGCCT 1560
      |||...|||
DB 2686 CAG...TCTGGCTCACTGCAATCCCGAGTCCCGAGGTTCAAGGATTCCTCGCTCAGTCT 2745

QY 1561 CCC...AGTAGTGGGACTACAGGCACTTCCCGAGGCTGGCTGCTCAAACTTTTCTGTCTATAATTG 1620
      |||...|||
```

```
Db 2746 CCCAAGCAGCTGGGATTACAGGTGCTGCCACCATGCCAGCTAAATTTTGTATTTTAG 2805
QY 1621 TAGAGATGGGATTTCACTTTTGGCCAAAGCTGCTCWAACATTTTGTGTCATATTTG 1680
Db 2806 TAGAGACAGGTTTCGCGGTGGCCAGGCTGCTCGAACT----- 2848
QY 1681 TTGTAACTATTGTCCTTTTCTGAGTAGGGCCCCCAGACCAAAATAAATCTTAG 1740
Db 2849 ----- 2873
QY 1741 AATCCAAATCAGTGTGTTGGTGAACACATGCTACTTTGAGAACACAGTGTGACAGGC 1800
Db 2874 TCGCCCTCCCAAAGTGTGGGATTACAGGTGTGAGCCACTGCCGCCAGCGGGTTTCATC 2933
QY 1801 CTCAGGAGTAGGTGATCTCTGCTCGAAGAGAAATAGAAATATATCTCCGGGCC 1860
Db 2934 CTTAATA-CATACATATAGATATAGATTCTGTTTATCTAAAGTCTTTATAAGGCC 2992
QY 1861 AGCGTGTGGCTCATGCCCTGTAATCCAGCACTTTGGGAGGCCAAGGCATGTGGATCAC 1920
Db 2993 GGGCGCGTGGCTCACGCCCTGTAATCCAGCACTTTGGGAGGCCGAGCGGGGATCAC 3052
QY 1921 CTGAGGTGAGAGTTCAAACACAGCTGGCCACATGGTGAACCCCGTCTCTACTAAA 1980
Db 3053 --GAGGTGAGAGTACAGACCATCCCGCTAAACAGTGAACCCCGTCTCTACTAAA 3110
QY 1981 ATACAAAATTAAGTAAAGTGTGGTGGCGCATGCTGTAATCCAGCTACTTTGGAGGCT 2040
Db 3111 ATACAAAATTAAGCGGGCGTAGTGGCGGGCGCTGTAGTCCAGCTACTTTGGAGGCT 3170
QY 2041 GAGCAGGAGAAATTTTGAACCGGGGAGGAGAGAGTGGTGCAGTGAAGCAGATCACACCA 2100
Db 3171 GAGCAGGAGAAATGGCTGAACCGGGGAGGAGAGTGGTGCAGTGAAGCAGATCCCGCCA 3230
QY 2101 CTGCATCCAGCTGGGGAGAGAGAGAGACTTCCTCTCAAAAACAAAACAAAAGA 2160
Db 3231 CTGCATCCAGCTGGGGAGAGAGAGACTTCCTCTCAAAAACAAAACAAAAGAAA 3290
QY 2161 ATTAA 2165
Db 3291 AAAAA 3295

RESULT 13
US-09-851-896-3
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEPENDENT)
; FILE REFERENCE: RTS-0220
; CURRENT APPLICATION NUMBER: US/09/851,896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-851-896-3

Query Match 7.2%; Score 297.2; DB 4; Length 70000;
Best Local Similarity 68.4%; Pred. No. 1.8e-62;
Matches 491; Conservative 0; Mismatches 208; Indels 19; Gaps 5;
```

```
QY 1441 TTTTATTTTATTTTGTGAGAAAGAGTCTGCTCTT-TCACCTGGCTGGAGGCGAGTGGT 1499
Db 49703 TTTATTTATTTTGTGAGACAGAGTTTCGCTTTTGTGGCCAGGCTGGAGTGCATGGC 49762
QY 1500 GTGATCTCAGCTCAGTCAACTTTGGCTCCCGGGTTTCAAGCAATTTCTCTGCTTCAGCC 1559
```

```
Db 49763 ACAGTCTGGCTACCGCAACCTCTGCTCTGGGTCAAGCATCTCTCCGCTCAGCT 49822
QY 1560 TCCCAAGTAGTGGGACTACAGGCACCTTCCACCATGCCCAGCTGATTTTGTATTTTAA 1619
Db 49823 TCCGAGTAGTGGGATTACGTTTCATGCGCACACACCTGGCT-AAATTTGTATTTTAA 49881
QY 1620 GTAGAGATGGGATTTCACTTTGTTGGCCAGCTGCTCAAACTTTTGTGCTGTCATATT 1679
Db 49882 GTAGAGACGGGGTTTCTCCATGTGTGTCAGGCTGGTCTTGAACCT-----CCTGACCT 49933
QY 1680 GTTGTAACTATTGTCCTTTTGTGAGTAGGGCCCCCAGACCAAAATAAATCTTA 1739
Db 49934 CAGGTGTCCACCTGCTCAGCTCCCAACTGCTGGGATTACAGCGTGACCCAGCTG 49993
QY 1740 GAATCCAAATCAGTGTGTTGGTTTGACCACTGTCACTTTGAGAACACACAGTGTGACAGGG 1799
Db 49994 CCGGGCTGGCTTAACTGACTTAAATTTATTTTGAAGAAATGTT-----G 50046
QY 1800 CCTCAGGAGTAGAGTGTCTCTGCTCGAAGAGAAATAGAAATATTTCTCCGGGC 1859
Db 50047 CATCACTACCATAAATGGCCATCCATCCAGTGCACCAACCAATATAAGGTAACTGC 50106
QY 1860 CAGCGCTGGTGCCTGTAAATCCAGCACTTTGGGAGGCCAAGGCATGTGGATCA 1919
Db 50107 AGGCGCGGTGGCTCAAGCTGTAATCCAGCACTTTGGGAGGCCGAGTGGCGAATCA 50166
QY 1920 CTTGAGTGTAGAGTGTCAAACACAGCTGGCCACATGGTGAACCCCGTCTCTACTAAA 1979
Db 50167 C--GAGTGTAGGAGTTCAAGACAGCAGCTGGCCACATGGTGAACCCCGTCTCTACTAAA 50224
QY 1980 AATACAAAATTAAGTAAAGTGTGGTGGCGCATGCTGTAATCCAGCTACTTTGGGAGGG 2039
Db 50225 AATACAAAATTAAGTGTGGCGTAGTGTGGCCACCTGTAATFACCAGCTACTCTGGAGGC 50284
QY 2040 TGAGGAGAGAGAAATTTTGAACCCCGGAGGAGAGTTCGCTGAAGCGAGATCACACC 2099
Db 50285 TGAGGAGAGAGAAATTTGAACTCGGTAGGTAGAGTTCGCTGAGTGTGAGTGCAGTGCAGCC 50344
QY 2100 ACTGCACTCCAGCTGGGGAGAGAGAGAGACTTCCTCTCAAAAACAAAACAAA 2157
Db 50345 ACTGCACTCCAGCTGGGTGACAGAGTGCAGCTCCGCTCTCAAAAACAAAACAAA 50402

RESULT 14
US-09-813-817-3
; Sequence 3, Application US/09813817
; Patent No. 6340583
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001178
; CURRENT APPLICATION NUMBER: US/09/813,817
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
US-09-813-817-3

Query Match 7.1%; Score 296.6; DB 4; Length 59065;
Best Local Similarity 68.2%; Pred. No. 2.3e-62;
Matches 518; Conservative 0; Mismatches 224; Indels 17; Gaps 7;
```

```
QY 1441 TTTTATTTTATTTTGTGAGAAAGAGTCTGCTCT-TCACCTGGCTGGAGGCGAGTGGT 1499
Db 11642 TTGTTTTTGTGTTTGTAGACAGTCTCACTCTGTTCCTCCAGGCTGGAGTGCATGGC 11701
QY 1500 GTGATCTCAGCTCAGTCAACTTTGCGCTCCCGGGTTTCAAGCAATTTCTCTGCTTCAGCC 1559
```

Qy	1441	TTTTTTTTTTTTTTTTTTTGAGAAAGAGCTGTGGCTCT--TTCACCCCTGGCTGAGGCGCAGTGGT	1499
Db	11642		
Db	11702	TGTGTTTGTGTTTGTATTTTATGACAGAGACTCACTCTGTCTGCCAAGGCTGGAGTGCAGTGGC	11703
Qy	1500	GTGATCTCAGCTCACTGCAACTTTTCGCTCCCGGCTTCAAGCAAATTCCTCCTGCTTCAGOC	1559
Db	11702		
Qy	1560	TCCCAAGTAGCTGGGACTACAGSCACTTCCCACCATTGCCAGCTGATTTTGTATTTTTTA	1619
Db	11762	TCCCAAGTAGCTGGGATTA-AGGCACGCGCCCCCATACCCASCATAATTTTATATTTTA	11820
Qy	1620	GTAGAGATGGGAATTCACATTTGTTGGCCAAGCTGGTCTCAAACTTTTGTCTGCTCAFAAT	1679
Db	11821		
Db	11821	GTAGAGATGGGGTTTGTGCATGTTGGCCAAGCTAGTCTCGAACTCCTGACCTCAAGTAGT	11880
Qy	1680	--GTGTAACATATTGTTTCCTTTTCTCAGGTGAGGCCCCCAGACAAAAAAAATAAATC	1736
Db	11881	CCACTGCTCGGCTCCCAAGTGCTGGGATTACAGGTGTGAGCCACTGCACCTGGCTG	11940
Qy	1737	TTAGAATCCAAATCAGTGTGTTTGTGACCACCTCACTTGAGAACACACACTGTGAOCA	1796
Db	11941	GAAGGAGTGATCTTAAAAAIAAAAAAAAAACAIAAAAAAAAAAACTTCACTGTGTCACTCTGTGT	12000
Qy	1797	GGGCTCAGGAGTAGAGGTGATC-----TCTGCTGAAAGAGNAATAGATGAAAT	1848
Db	12001	GTCTCTCTCACTTGTATATCTATCTCCAACTTCCCCAGTGTCTTGGATAAAGACCACCAATC	12060
Qy	1849	ATTCTCCGGGCGAGCGGTGGTGGCTCATGCCCTGAATCCAGCACTTTGGGAGGCCAAG	1908
Db	12061	CTTAACTTGGCCAGCGCGGTGGCTCACACCTATCATCTCAGCACTTTGGGAGGCCGAG	12120
Qy	1909	CATGTGGATCAOCTGAGTCAAGGATTCAAACACCAAGCTGGCCACATGGTGAACCCCG	1968
Db	12121	CAGGCAGATCA--TGAAGTCAAGAGATTGAGACCATCTCTGGCCACATGGTGAACCCCA	12170
Qy	1969	TCTCTACTAAAAATACAAAAATTAAGCTAAGTGTGGTGGCGCATGCTGTATATCCAGCT	2028
Db	12179	TCTCTACTAAAAATAC-AAAAATTAAGTGTGGTGGCGGTGTGCCTGTAGTCCAGCT	12233
Qy	2029	ACTTGGGAGSGTGAGCGAGGAGAAATTTCTTGAACCGGGAGCGAGGTTGCAGTGAAC	2088
Db	12238	ACTTGGGAGSGTGAGCGAGGAGAAATCACTTGAACCTGGGAGCGAGGTTGCAGTGAAC	12297
Qy	2089	GAGATCACACCACTGCATCCAGCTCGGGGAGAGCGAGACTTCTCTCTCAAAAAACA	2148
Db	12298	CAGATCAGCCACTGCATCTCCAGCT--GGTGACAGTAAGACTCCATCTCAAAAAAAA	12356
Qy	2149	AAAAACAAAGAAATTAAGCAATTTAGACATTCGAGAG	2187
Db	12357	AAAAAAAAAAAAAAAAATTCCTTAATTTGGCTACAGTAGAG	12395

Search completed: May 20, 2003, 19:56:27  
Job time : 3074 secs

Query Match 7.1%; Score 296.6; DB 4; Length 59065;  
Best Local Similarity 68.2%;  
Pred. No. 2.3e-62;  
Matches 518; Conservative 0; Mismatches 224; Indels 17;



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 19:06:28 ; Search time 489 Seconds  
(without alignments)  
11206.377 Million cell updates/sec

Title: US-10-016-725-15  
Perfect score: 4150  
Sequence: 1 aggaattcatcatttaaat.....tgagttccagcctggccc 4150

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 828747 seqs, 560231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4150	100.0	12	US-10-016-725-15	Sequence 15, Appl
C 2	429	10.3	17704	9 US-10-092-154-1865	Sequence 1865, Ap
C 3	429	10.3	17704	10 US-09-764-847-1865	Sequence 1865, Ap
C 4	427.4	10.3	17707	9 US-10-092-154-1864	Sequence 1864, Ap
C 5	427.4	10.3	17707	10 US-09-764-847-1864	Sequence 1864, Ap
C 6	409.8	9.9	5648	9 US-09-981-353-183	Sequence 183, Ap
C 7	409.8	9.9	5648	10 US-09-919-172-70	Sequence 70, Appl
C 8	349	8.4	13216	9 US-09-764-891-7810	Sequence 7810, Ap
C 9	347.4	8.4	29163	9 US-09-764-891-7809	Sequence 7809, Ap
C 10	342.8	8.3	26928	9 US-10-020-141-7	Sequence 7, Appl
C 11	342.8	8.3	26928	10 US-09-880-107-2278	Sequence 2278, Ap
C 12	329.6	7.9	9372	10 US-09-764-877-3049	Sequence 3049, Ap
C 13	329.6	7.9	9372	10 US-09-764-877-3244	Sequence 3244, Ap
C 14	327.6	7.9	17397	9 US-10-091-504-1945	Sequence 1945, Ap
C 15	327.6	7.9	17397	10 US-09-764-869-1945	Sequence 1945, Ap
C 16	327.6	7.9	19334	9 US-10-091-504-1943	Sequence 1943, Ap
C 17	327.6	7.9	19334	10 US-09-764-869-1943	Sequence 1943, Ap
C 18	327.6	7.9	19345	9 US-10-091-504-1944	Sequence 1944, Ap
C 19	327.6	7.9	19345	10 US-09-764-869-1944	Sequence 1944, Ap

```

c 20 327.6 7.9 25423 9 US-10-072-349-193 Sequence 193, App
c 21 327.6 7.9 25423 10 US-09-764-855-193 Sequence 193, App
c 22 327 7.9 5932 9 US-09-764-891-6396 Sequence 6396, Ap
c 23 326.4 7.9 32148 9 US-09-764-891-6906 Sequence 6906, Ap
c 24 326 7.9 25424 9 US-10-072-349-194 Sequence 194, App
c 25 326 7.9 25424 10 US-09-764-855-194 Sequence 194, App
c 26 325.4 7.8 32145 9 US-09-764-891-7319 Sequence 7319, Ap
c 27 324.4 7.8 32191 9 US-09-764-891-6304 Sequence 6304, Ap
c 28 324 7.8 3471 9 US-10-092-154-1893 Sequence 1893, Ap
c 29 324 7.8 3471 10 US-09-764-847-1893 Sequence 1893, Ap
c 30 323.2 7.8 32134 9 US-09-764-891-6303 Sequence 6303, Ap
c 31 321.8 7.8 32216 9 US-09-764-891-9613 Sequence 9613, Ap
c 32 320.4 7.7 465237 10 US-09-933-267A-1 Sequence 1, Appl
c 33 320.2 7.7 7703 9 US-10-073-961-415 Sequence 415, App
c 34 320.2 7.7 7703 9 US-10-073-961-415 Sequence 415, App
c 35 320.2 7.7 7703 10 US-09-764-887-415 Sequence 415, App
c 36 320.2 7.7 7703 10 US-09-764-887-456 Sequence 456, App
c 37 319.6 7.7 32192 9 US-10-092-154-1416 Sequence 1416, Ap
c 38 319.6 7.7 32192 9 US-09-764-891-7945 Sequence 7945, Ap
c 39 319.6 7.7 32192 9 US-09-764-891-8220 Sequence 8220, Ap
c 40 319.6 7.7 32192 9 US-09-764-891-8319 Sequence 8319, Ap
c 41 319.6 7.7 32192 10 US-09-764-847-1416 Sequence 1416, Ap
c 42 319.6 7.7 65608 9 US-09-954-531-180 Sequence 180, App
c 43 319.6 7.7 65608 10 US-09-962-436-292 Sequence 292, App
c 44 319.6 7.7 65608 10 US-09-962-832-119 Sequence 119, App
c 45 317.8 7.7 10527 9 US-10-116-420-100 Sequence 100, App

```

## ALIGNMENTS

```

RESULT 1
US-10-016-725-15
; Sequence 15, Application US/10016725
; Patent No. US20020151018A1
; GENERAL INFORMATION:
; APPLICANT: Protuity, Stephen
; APPLICANT: Zhang, Lin
; APPLICANT: Stenn, Kurt
; TITLE OF INVENTION: Stearyl-CoA Desaturase Promoter
; FILE REFERENCE: J6J2065
; CURRENT APPLICATION NUMBER: US/10/016,725
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 15
; LENGTH: 4150
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-725-15

```

```

Query Match 100.0%; Score 4150; DB 12; Length 4150;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAATTCATCCATTTAAATCATACAAATTTAATGGCTTTTAGTATATATTCACAGTTTGC 60
Db 1 AGGAATTCATCCATTTAAATCATACAAATTTAATGGCTTTTAGTATATATTCACAGTTTGC 60
QY 61 ATCCATCACAAATCCATTTTAAAGACAGTTTATTTACTCCAAAATAAACCCTGCATTCCTT 120
Db 61 ATCCATCACAAATCCATTTTAAAGACAGTTTATTTACTCCAAAATAAACCCTGCATTCCTT 120
QY 121 AGGCATCACCCCCCAACATCCTCCATCCTCTCCAGCCCTGGCAACCAACCAATCTAC 180
Db 121 AGGCATCACCCCCCAACATCCTCCATCCTCTCCAGCCCTGGCAACCAACCAATCTAC 180
QY 181 TTCTGTCTCTATAAAATTTGCCAATTTCTGCACATTTCTGACATTTCTAATAATGAAGCAACAACAT 240
Db 181 TTCTGTCTCTATAAAATTTGCCAATTTCTGCACATTTCTGACATTTCTAATAATGAAGCAACAACAT 240
QY 241 GTGAGACTTTGTGACAGCTGGCTTCCTTACCTAGCATTCATTTTAAAGCTCATTAATCTTA 300

```

Db 241 GTGAGACTTTGTGACTGGCTGCTTTCACCTTAGCAATCTATTTTAAAGGCTCATATGTTA 300  
QY 301 CAGTACTAGCAGTACTTTCATCTTTTTTATCTCAAAAGGTATCCACTGTGGGGTAT 360  
Db 301 CAGTACTAGCAGTACTTTCATCTTTTTTATCTCAAAAGGTATCCACTGTGGGGTAT 360  
QY 361 CCATATCATATTTATAGACAGAGTCTCAGTGTGACCCAGGCTGGAGTGCAGTGGC 420  
Db 361 CCATATCATATTTATAGACAGAGTCTCAGTGTGACCCAGGCTGGAGTGCAGTGGC 420  
QY 421 ACATATAGCTACTGTAACTCAAACTCCTGGGCTCAAGTGAATCTACTACCTCAGCC 480  
Db 421 ACATATAGCTACTGTAACTCAAACTCCTGGGCTCAAGTGAATCTACTACCTCAGCC 480  
QY 481 TCCAGAGTAGCTAGGACTACAGGCACACAGCCATACCTGGCTAAATTTTTTTTTTAAT 540  
Db 481 TCCAGAGTAGCTAGGACTACAGGCACACAGCCATACCTGGCTAAATTTTTTTTTTAAT 540  
QY 541 TTTCATTTTATGTATTCATTTCTTTCTTTTGTGTTTGTGTTTGTGATAGAGTCTC 600  
Db 541 TTTCATTTTATGTATTCATTTCTTTCTTTTGTGTTTGTGTTTGTGATAGAGTCTC 600  
QY 601 ACTTTGTACCCAGGCTGGAGGAGTGGCATGGTGTGACAGCTGAGAGCCTTGACTTCCT 660  
Db 601 ACTTTGTACCCAGGCTGGAGGAGTGGCATGGTGTGACAGCTGAGAGCCTTGACTTCCT 660  
QY 661 GGGCTCAAGTGATCCTCCTGCTCAGCCTCCCAAGTAGCTGGGACTACAAACAGCTGTCA 720  
Db 661 GGGCTCAAGTGATCCTCCTGCTCAGCCTCCCAAGTAGCTGGGACTACAAACAGCTGTCA 720  
QY 721 CCATGCTGGCTGATATTTTTTCTTTGAAACAGGGTATCAGTCTGTGCCCAGGCTGGA 780  
Db 721 CCATGCTGGCTGATATTTTTTCTTTGAAACAGGGTATCAGTCTGTGCCCAGGCTGGA 780  
QY 781 GTACAGTGGGTAATAATAGCTCAGCTCAGCCTCCCTCCTGGCTCAAGCAATCCGCTG 840  
Db 781 GTACAGTGGGTAATAATAGCTCAGCTCAGCCTCCCTCCTGGCTCAAGCAATCCGCTG 840  
QY 841 GCCTCAGCATCCTGAGTAGCTGGGACTACAGGCTTGTGCCACAGGCCAGCTAAAGTTT 900  
Db 841 GCCTCAGCATCCTGAGTAGCTGGGACTACAGGCTTGTGCCACAGGCCAGCTAAAGTTT 900  
QY 901 AAAAAATGATTTTGTATAGAGAGGCTTGTATGTTGCTCAGGCTGTATTTTATG 960  
Db 901 AAAAAATGATTTTGTATAGAGAGGCTTGTATGTTGCTCAGGCTGTATTTTATG 960  
QY 961 TTGAGACAAGGCTCAGTATGTTGCCATGATCCGCCACCTCCACTTCCCAAAAGTGCTCA 1020  
Db 961 TTGAGACAAGGCTCAGTATGTTGCCATGATCCGCCACCTCCACTTCCCAAAAGTGCTCA 1020  
QY 1021 TCTTATCTGTTCAATAGTCAAGTTGACAGACATTTAGTGTGTTTCCACTTTTTCACCATTA 1080  
Db 1021 TCTTATCTGTTCAATAGTCAAGTTGACAGACATTTAGTGTGTTTCCACTTTTTCACCATTA 1080  
QY 1081 TGAATATCTCCAGTGAATATTCATGTATACATTTGTGTTGCGCATATGTTTCAATTC 1140  
Db 1081 TGAATATCTCCAGTGAATATTCATGTATACATTTGTGTTGCGCATATGTTTCAATTC 1140  
QY 1141 GTTGGGTTTATATCTAGAGTGGAAATGCTGGATCCCGGTAATATTTGACAGGCAGAG 1200  
Db 1141 GTTGGGTTTATATCTAGAGTGGAAATGCTGGATCCCGGTAATATTTGACAGGCAGAG 1200  
QY 1201 TTCAGGGGAAGAAAACCTTGGGAAAATGAAGCATGTTTGAATAATCAGCAAGTGCAGGG 1260  
Db 1201 TTCAGGGGAAGAAAACCTTGGGAAAATGAAGCATGTTTGAATAATCAGCAAGTGCAGGG 1260  
QY 1261 GTTTTTCGGAGTTTATTTTATATCTGTTGACAAATGTGCAATTTGATGAAGATACAAG 1320  
Db 1261 GTTTTTCGGAGTTTATTTTATATCTGTTGACAAATGTGCAATTTGATGAAGATACAAG 1320  
QY 1321 TTATACTAAGTGAAGTGAATTAAGGCTGGAATAGGCGCTTCAGAGTAAAATCATGA 1380  
Db 1321 TTATACTAAGTGAAGTGAATTAAGGCTGGAATAGGCGCTTCAGAGTAAAATCATGA 1380

QY 1381 AGCACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAAAATAATAAAAAATCAAA 1440  
Db 1381 AGCACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAAAATAATAAAAAATCAAA 1440  
QY 1441 TTTTTTTTTTTTTTTTCAGAAAGAGTCTGCTCTTTTCCCTGCTGCTGAGGGCAGTGGTG 1500  
Db 1441 TTTTTTTTTTTTTTTTCAGAAAGAGTCTGCTCTTTTCCCTGCTGCTGAGGGCAGTGGTG 1500  
QY 1501 TGATCTCAGCTCACTGCAAACTTTTCGCTCCCGGGTTTCAAGCAATTCCTCTGCTCAGCCT 1560  
Db 1501 TGATCTCAGCTCACTGCAAACTTTTCGCTCCCGGGTTTCAAGCAATTCCTCTGCTCAGCCT 1560  
QY 1561 CCCAAGTAGCTGGGACTACAGGCATTTCCCAACCATGCCCAGCTGATTTTGTATTTTAG 1620  
Db 1561 CCCAAGTAGCTGGGACTACAGGCATTTCCCAACCATGCCCAGCTGATTTTGTATTTTAG 1620  
QY 1621 TAGAGATGGGATTTCACTTTTGTGGCAAGCTGCTCAAACCTTTTGTCTGCATATAATTG 1680  
Db 1621 TAGAGATGGGATTTCACTTTTGTGGCAAGCTGCTCAAACCTTTTGTCTGCATATAATTG 1680  
QY 1681 TTGTAACATATTTTCCTTTTGTGCTGAGGTAGGGCCCCAGACCAAAAAATAAATCCTTAG 1740  
Db 1681 TTGTAACATATTTTCCTTTTGTGCTGAGGTAGGGCCCCAGACCAAAAAATAAATCCTTAG 1740  
QY 1741 AATCCAAATCAGTGTGTTTGTGACCACTGCTCACTTTGAGAACCCACAGTGTGACCAGGGC 1800  
Db 1741 AATCCAAATCAGTGTGTTTGTGACCACTGCTCACTTTGAGAACCCACAGTGTGACCAGGGC 1800  
QY 1801 CTCAGGAGTAGAGTGAATCTCTGCTCGAAAAGAGAAATAGAAATGAAATATTTCTCCGGGCC 1860  
Db 1801 CTCAGGAGTAGAGTGAATCTCTGCTCGAAAAGAGAAATAGAAATGAAATATTTCTCCGGGCC 1860  
QY 1861 AGCGTGGTGGCTCATGCCCTGTAATCCAGCAGCTTTGGAGGCCCAAGCATGTGGATCAC 1920  
Db 1861 AGCGTGGTGGCTCATGCCCTGTAATCCAGCAGCTTTGGAGGCCCAAGCATGTGGATCAC 1920  
QY 1921 CTGAGGTTCAGAGGTTCAAAACCCAGCCTGGCCAAACATGCTGAAACCCCGCTCTACTAAAA 1980  
Db 1921 CTGAGGTTCAGAGGTTCAAAACCCAGCCTGGCCAAACATGCTGAAACCCCGCTCTACTAAAA 1980  
QY 1981 ATACAAAAATTAAGTGAAGTGTGGTGGCGATGCTGTAATCCAGCTACTTGGGAGGGT 2040  
Db 1981 ATACAAAAATTAAGTGAAGTGTGGTGGCGATGCTGTAATCCAGCTACTTGGGAGGGT 2040  
QY 2041 GAGCAGGAGAAATTTCTTGAACCCGGGAGGAGGTTGAGTGAAGCAGATACACACCA 2100  
Db 2041 GAGCAGGAGAAATTTCTTGAACCCGGGAGGAGGTTGAGTGAAGCAGATACACACCA 2100  
QY 2101 CTGCACCTCCAGCCTGGGGGAGAGCGAGACTTCTCTCAAAAAACAAAAACAAAAAGA 2160  
Db 2101 CTGCACCTCCAGCCTGGGGGAGAGCGAGACTTCTCTCAAAAAACAAAAACAAAAAGA 2160  
QY 2161 ATTAAGCAAAATTAAGCATTCAGAGAGAACCTGAAGGGGTCAGACCATACAGATTC 2220  
Db 2161 ATTAAGCAAAATTAAGCATTCAGAGAGAACCTGAAGGGGTCAGACCATACAGATTC 2220  
QY 2221 TGTGCCACATGCCAAGTACTTCTGAGGCATGACTGGATGAGCTTCCACATCTGAATCA 2280  
Db 2221 TGTGCCACATGCCAAGTACTTCTGAGGCATGACTGGATGAGCTTCCACATCTGAATCA 2280  
QY 2281 TCCAGTCTTGTTCAGAACTTTTCACCGGACAGGGAGCCAGGACTGGAATGCAAGTCTCCT 2340  
Db 2281 TCCAGTCTTGTTCAGAACTTTTCACCGGACAGGGAGCCAGGACTGGAATGCAAGTCTCCT 2340  
QY 2341 GGTCACTGGCCAGAGAGTTGGCCTTGACCCCTGAGACCAAGTGGCCAAACAAAGGAGCTGCT 2400  
Db 2341 GGTCACTGGCCAGAGAGTTGGCCTTGACCCCTGAGACCAAGTGGCCAAACAAAGGAGCTGCT 2400  
QY 2401 AGTCTACCTCCAGAGAAATCCCAAGTGTCTTCTTCCGGAAGTGAATCATTTGGGCGCAG 2460  
Db 2401 AGTCTACCTCCAGAGAAATCCCAAGTGTCTTCTTCCGGAAGTGAATCATTTGGGCGCAG 2460

Qy	2461	CAC	TCCGTAATTTTCTCCTCTCCAGGGGAAGATCCTTAGGCAGTAATTGGGAAGAACA	2520
Db	2461	CAC	TCCGTAATTTTCTCCTCTTCGCCAGGGGAAGATCCTTAGGCAGTAATTGGGAAGAACA	2520
Qy	2521	TGG	CATGGAAGCACACCGSGTGAAATGATACCTGCCTGTCTGAGCTCTCATGTAA	2580
Db	2521	TGG	CATGGAAGCACACCGGGTGAATGCATACCTGCCTGTCTGAGCTCTCATGTAA	2580
Qy	2581	GGT	CTCTACAGACACGAAAGATGGGSCACAGGACAGATCAGTAGGTCAGAGCATC	2640
Db	2581	GGCT	CTCTACAGACACGAAAGATGGGSCACAGGACAGATCAGTAGGTCAGAGCATC	2640
Qy	2641	TCAGG	ACCAGGGCAATATGTCTTGACAGGATTAAGACTTGGCTCTCATATGGT	2700
Db	2641	TCAGG	ACCAGGGCAATATGTCTTGACAGGATTAAGACTTGGCTCTCATATGGT	2700
Qy	2701	GTTT	CTGGCTCAACTGCCAGTCCGTCACITACTCTGTGTGACCATGGCAAGTTA	2760
Db	2701	GTTT	CTGGCTCAACTGCCAGTCCGTCACITACTCTGTGTGACCATGGCAAGTTA	2760
Qy	2761	TTCCAT	CTCTCOATNTCTTTTCTCACHTTTAAAATGGAATAATGGGTACCACTCC	2820
Db	2761	TTCCAT	CTCTCOATNTCTTTTCTCACHTTTAAAATGGAATAATGGGTACCACTCC	2820
Qy	2821	CAGG	GTACAGAGGCTTACAAAAAGATTCTTGTGAATTGGCTGCAGTAATAATTC	2880
Db	2821	CAGG	GTACAGAGGCTTACAAAAAGATTCTTGTGAATTGGCTGCAGTAATAATTC	2880
Qy	2881	AATAC	TGCCAGCTATTCTATTTCACATPCAAAGCCCTTTCGCTGTGTGGGTGAAAA	2940
Db	2881	AATAC	TGCCAGCTATTCTATTTCACATPCAAAGCCCTTTCGCTGTGTGGGTGAAAA	2940
Qy	2941	CACAT	GTCACTGTTTCTCAGCGTTTCCAAGAAGATTCCAANAATTAACAACCTGCCAG	3000
Db	2941	CACAT	GTCACTGTTTCTCAGCGTTTCCAAGAAGATTCCAANAATTAACAACCTGCCAG	3000
Qy	3001	TCGAGA	ATCTCCAAACATCCGACGAPCTCTGAGGCGCGGCTTGGGGATGGGAC	3060
Db	3001	TCGAGA	ATCTCCAAACATCCGACGAPCTCTGAGGCGCGGCTTGGGGATGGGAC	3060
Qy	3061	TGCG	CCCGSGPCTGAACAGSATCGTGCAGCAGCACACACACAGCCAGCTG	3120
Db	3061	TGCG	CCCGSGPCTGAACAGSATCGTGCAGCAGCACACACACAGCCAGCTG	3120
Qy	3121	TGTGTG	CGCGGATCCGCTCCCGGTGAGACGCGTGGCTGTGSGCGGGGC	3180
Db	3121	TGTGTG	CGCGGATCCGCTCCCGGTGAGACGCGTGGCTGTGSGCGGGGC	3180
Qy	3181	AGAG	CCATTGTCGAGGCGTACCGAGCCCCCGCGCTCGCCCGGAGGAGCGGGCT	3240
Db	3181	AGAG	CCATTGTCGAGGCGTACCGAGCCCCCGCGCTCGCCCGGAGGAGCGGGCT	3240
Qy	3241	TCCG	CGTCCCAGCTCCAGATCTCTGGSGTGSCTGCCACGCTCTCCCTGCCACGCGCTG	3300
Db	3241	TCCG	CGTCCCAGCTCCAGATCTCTGGSGTGSCTGCCACGCTCTCCCTGCCACGCGCTG	3300
Qy	3301	GGGG	ACGGAACGCGGAGATGTATGTGTGGCGCCCCCGAGGGTTCCACCACT	3360
Db	3301	GGGG	ACGGAACGCGGAGATGTATGTGTGGCGCCCCCGAGGGTTCCACCACT	3360
Qy	3361	GTTT	CTTGAAACTTCCCAGTGGCCCAACCCGTTCTCCGTGTGCCGAGGGCGGT	3420
Db	3361	GTTT	CTTGAGAACTTCCCAGTGGCCCAACCCGTTCTCCGTGTGCCGAGGGCGGT	3420
Qy	3421	CCTG	GGCTAGGCTCCGCGCCCCAGCCCCAAACCGGGTCCCCAGCCCTTCCAGAGAAA	3480
Db	3421	CCTG	GGCTAGGCTCCGCGCCCCAGCCCCAAACCGGGTCCCCAGCCCTTCCAGAGAAA	3480
Qy	3481	GCTC	CCAGCGGGATTCGCGGCAGAGGCCAGCGCGGTGGAAGAGAACCTCAGAGG	3540
Db	3481	GCTC	CCAGCGGGATTCGCGGCAGAGGCCAGCGCGGTGGAAGAGAACCTCAGAGG	3540
Qy	3541	AGAA	CAGAGGGAGGGGAGGAGCTGGCGCAGAGGGAACAGCAGATTGCGCGCA	3600

D	b		3541	AGAAACAGAGGGGAGGGGAGCGAGGAGCTGGCGGCAGAGGAACAAGCAGATTGGCCGCA	3600
Q	y		3601	GCCAATGGCAAGCGGACGAGGAGGTGGCCACCATAATTCCTTCGGCCCAATCACGAGCCGGA	3660
D	b		3601	GCCAATGGCAAGCGGACGAGGAGGTGGCCACCATAATTCCTTCGGCCCAATCACGAGCCGGA	3660
Q	y		3661	GTTTACAGAAGCCTCATTAGCATTTCCCAGAGGCGACAGGGGACAGAGGCCGGGTGG	3720
D	b		3661	GTTTACAGAAGCCTCATTAGCATTTCCCAGAGGCGACAGGGGACAGAGGCCGGGTGG	3720
Q	y		3721	TGTGTGTGCGGTGTGCGCAGCATPCCCGGCGCCTGCTGTGGGTTCGCCGAGCCTCGGCC	3780
D	b		3721	TGTGTGTGCGGTGTGCGCAGCATPCCCGGCGCCTGCTGTGGGTTCGCCGAGCCTCGGCC	3780
Q	y		3781	TGTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCGCGCAGTCAAATCC	3840
D	b		3781	TGTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCGCGCAGTCAAATCC	3840
Q	y		3841	TCGCACTTTGCCCTGCTTTGGCAGCGGATAAAGGGGGTGAGGAAATACCGGACACGGT	3900
D	b		3841	TCGCACTTTGCCCTGCTTTGGCAGCGGATAAAGGGGGTGAGGAAATACCGGACACGGT	3900
Q	y		3901	CACCGGTTGCCAGCTTAGCTTTAAATTCGCGGCTCGGGGACTCCACGCACCGCGGCT	3960
D	b		3901	CACCGGTTGCCAGCTTAGCTTTAAATTCGCGGCTCGGGGACTCCACGCACCGCGGCT	3960
Q	y		3961	AGCGCGCAACACGCTAGCGTGCAAGGGCGCGCGGCTCAGCGCTACCGCGCGGGCTTCG	4020
D	b		3961	AGCGCGCAACACGCTAGCGTGCAAGGGCGCGCGGCTCAGCGCTACCGCGCGGGCTTCG	4020
Q	y		4021	AAACCGCAGTCTCCGGGACCCCGAACCTCCGCTCCGGAGCCTCAGCCCCCTGGAAGTG	4080
D	b		4021	AAACCGCAGTCTCCGGGACCCCGAACCTCCGCTCCGGAGCCTCAGCCCCCTGGAAGTG	4080
Q	y		4081	ATCCCGGCATCCGAGAGCCAAAGATGCCGCCCACTTGTCTCAGGACGATGTGAGTTCCC	4140
D	b		4081	ATCCCGGCATCCGAGAGCCAAAGATGCCGCCCACTTGTCTCAGGACGATGTGAGTTCCC	4140
Q	y		4141	AGCCTGGCCC	4150
D	b		4141	AGCCTGGCCC	4150
<b>RESULT 2</b>					
US-10-092-154-1865/c					
; Sequence 1865, Application US/10092154					
; Publication NO. US20030054375A1					
; GENERAL INFORMATION:					
; APPLICANT: Rosen et al.					
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies					
; FILE REFERENCE: PC009CI					
; CURRENT APPLICATION NUMBER: US/10/092,154					
; CURRENT FILING DATE: 2002-03-07					
; NUMBER OF SEQ ID NOS: 2003					
; Prior Application removed - See File Wrapper or Palm					
; SOFTWARE: PatentIn Ver. 2.0					
; SEQ ID NO 1865					
; LENGTH: 17704					
; TYPE: DNA					
; ORGANISM: Homo sapiens					
US-10-092-154-1865					

## RESULTS

```

US-10-092-154-1865/c
; Sequence 1865, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1865
; LENGTH: 17704
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1865

```

	Query Match	10.3%	Score 429;	DB 9;	Length 17704;
	Best Local Similarity	100.0%;	Pred. No. 4.4e-104;		
	Matches 429;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps
QY	3722	GTGGTGCGGTGTCGCGAGCATCCCGGGCCCTGCTCGGTGCGCGGAGCCTCGGCCT	3781		
Db	17704	GTGGTGCGGTGTCGCGAGCATCCCGGGCCCTGCTCGGTGCGCGGAGCCTCGGCCT			
QY	3782	CTGTCTCTCTCCCTCCCGCCCTTACTCTCAGCGGGACCGCGCGGCCCACTCTCT	3841		

Db 17644 CTGTCCTCCCTCCCGCCCTTACCTCCACGCGGACCGCCCGCGCAGTCAACTCCT 17585  
QY 3842 CGCACTTTGCCCTCGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 3901  
Db 17584 CGCACTTTGCCCTCGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 17525  
QY 3902 ACCGTTTGGCAGCTCTAGCCCTTTAAATTCGCGGCTCGGGAGCCTCCACGACCGGGCTA 3961  
Db 17524 ACCGTTTGGCAGCTCTAGCCCTTTAAATTCGCGGCTCGGGAGCCTCCACGACCGGGCTA 17465  
QY 3962 GCGCGGACAAACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 4021  
Db 17464 GCGCGGACAAACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 17405  
QY 4022 AACCGCAGTCTCCGCGGACCGCCGAACTCCGCTCCGAGCCTCAGCCCTCGGAAAGTGA 4081  
Db 17404 AACCGCAGTCTCCGCGGACCGCCGAACTCCGCTCCGAGCCTCAGCCCTCGGAAAGTGA 17345  
QY 4082 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTCTGAGGACGATGTAGTTTCCCA 4141  
Db 17344 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTCTGAGGACGATGTAGTTTCCCA 17285  
QY 4142 GCCTGGGCC 4150  
Db 17284 GCCTGGGCC 17276

## RESULT 3

US-09-764-847-1865/c  
; Sequence 1865, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764,847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1865  
; LENGTH: 17704  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-847-1865

Query Match 10.3%; Score 429; DB 10; Length 17704;  
Best Local Similarity 100.0%; Pred. No. 4.4e-104;  
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3722 GTGGTGTGCGGTGTCGGCAGCATCCCGGGCCCTGTGCGGTGCGCGGACCTTCGGCCT 3781  
Db 17704 GTGGTGTGCGGTGTCGGCAGCATCCCGGGCCCTGTGCGGTGCGCGGACCTTCGGCCT 17645  
QY 3782 CTGTCCTCTCCCTCCCGCCCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCCT 3841  
Db 17644 CTGTCCTCTCCCTCCCGCCCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCCT 17585  
QY 3842 CGCACTTTGCCCTCGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 3901  
Db 17584 CGCACTTTGCCCTCGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 17525  
QY 3902 ACCGTTTGGCAGCTCTAGCCCTTTAAATTCGCGGCTCGGGAGCCTCCACGACCGGGCTA 3961  
Db 17524 ACCGTTTGGCAGCTCTAGCCCTTTAAATTCGCGGCTCGGGAGCCTCCACGACCGGGCTA 17465  
QY 3962 GCGCGGACAAACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 4021  
Db 17464 GCGCGGACAAACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 17405  
QY 4022 AACCGCAGTCTCCGCGGACCGCCGAACTCCGCTCCGAGCCTCAGCCCTCGGAAAGTGA 4081  
Db 17404 AACCGCAGTCTCCGCGGACCGCCGAACTCCGCTCCGAGCCTCAGCCCTCGGAAAGTGA 17345

QY 4082 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTCTGAGGACGATGTAGTTTCCCA 4141  
Db 17344 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTCTGAGGACGATGTAGTTTCCCA 17285  
QY 4142 GCCTGGGCC 4150  
Db 17284 GCCTGGGCC 17276

## RESULT 4

US-10-032-154-1864/c  
; Sequence 1864, Application US/10092154  
; Publication No. US20030054375A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009C1  
; CURRENT APPLICATION NUMBER: US/10/092,154  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2003  
; Prior application removed - See File Wrapper or Palm  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1864  
; LENGTH: 17707  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-032-154-1864

Query Match 10.3%; Score 427.4; DB 9; Length 17707;  
Best Local Similarity 99.8%; Pred. No. 1.2e-103;  
Matches 428; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3722 GTGGTGTGCGGTGTCGGCAGCATCCCGGGCCCTGTGCGGTGCGCGGACCTTCGGCCT 3781  
Db 17707 GTGGTGTGCGGTGTCGGCAGCATCCCGGGCCCTGTGCGGTGCGCGGACCTTCGGCCT 17648  
QY 3782 CTGTCCTCTCCCTCCCGCCCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCCT 3841  
Db 17647 CTGTCCTCTCCCTCCCGCCCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCCT 17588  
QY 3842 CGCACTTTGCCCTCGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 3901  
Db 17587 CGCACTTTGCCCTCGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 17528  
QY 3902 ACCGTTTGGCAGCTCTAGCCCTTTAAATTCGCGGCTCGGGAGCCTCCACGACCGGGCTA 3961  
Db 17527 ACCGTTTGGCAGCTCTAGCCCTTTAAATTCGCGGCTCGGGAGCCTCCACGACCGGGCTA 17468  
QY 3962 GCGCGGACAAACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 4021  
Db 17467 GCGCGGACAAACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 17408  
QY 4022 AACCGCAGTCTCCGCGGACCGCCGAACTCCGCTCCGAGCCTCAGCCCTCGGAAAGTGA 4081  
Db 17407 AACCGCAGTCTCCGCGGACCGCCGAACTCCGCTCCGAGCCTCAGCCCTCGGAAAGTGA 17348  
QY 4082 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTCTGAGGACGATGTAGTTTCCCA 4141  
Db 17347 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTCTGAGGACGATGTAGTTTCCCA 17288  
QY 4142 GCCTGGGCC 4150  
Db 17287 GCCTGGGCC 17279

## RESULT 5

US-09-764-847-1864/c  
; Sequence 1864, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1864
; LENGTH: 17707
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-847-1864

Query Match          10.3%; Score 427.4; DB 10; Length 17707;
Best Local Similarity 99.5%; Pred. No. 1.2e-103;
Matches 428; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3722 GTGGTGTGCGTTCGCGACATCCCGCGCCCTTACCTCCAGCGGGACCGCCGCGGCTCGCGCTCGGCGCT 3781
Db 17707 GTGGTGTGCGTTCGCGACATCCCGCGCCCTTACCTCCAGCGGGACCGCCGCGGCTCGCGCTCGGCGCT 17648
QY 3782 CTGTCTCCCTCCCGCTCCCGCCCTTACCTCCAGCGGGACCGCCGCGGCTCGCGCTCGGCGCTCGGCGCT 3841
Db 17647 CTGTCTCCCTCCCGCTCCCGCCCTTACCTCCAGCGGGACCGCCGCGGCTCGCGCTCGGCGCTCGGCGCT 17588
QY 3842 CGCACTTTGCCCTCTGTTGGACGGGATTAAGGGGGCTGAGGAATACCGACACGGTC 3901
Db 17587 CGCACTTTGCCCTCTGTTGGACGGGATTAAGGGGGCTGAGGAATACCGACACGGTC 17528
QY 3902 ACCCGTTCAGCTCTAGCTTTAAATTCGGCTCGGGACCTCCACGACCGCGGCTA 3961
Db 17527 ACCCGTTCAGCTCTAGCTTTAAATTCGGCTCGGGACCTCCACGACCGCGGCTA 17468
QY 3962 GCGCCGACACACAGCTAGCTGCAAGCGCGCGGCTACGCGGTACCGCGGCGCTTCGA 4021
Db 17467 GCGCCGACACACAGCTAGCTGCAAGCGCGCGGCTACGCGGTACCGCGGCGCTTCGA 17408
QY 4022 AACCAGCTCTCCCGGACCGGAACTCGGCTCCGGAGGCTCAGCCCGCTGGAAGTGA 4081
Db 17407 AACCAGCTCTCCCGGACCGGAACTCGGCTCCGGAGGCTCAGCCCGCTGGAAGTGA 17348
QY 4082 TCCCGCATCCGAGAGCAAGATGCGCGGCCACTTGTCTGAGGACGATGTGAGTTTCCCA 4141
Db 17347 TCCCGCATCCGAGAGCAAGATGCGCGGCCACTTGTCTGAGGACGATGTGAGTTTCCCA 17288
QY 4142 GCGTGGCCC 4150
Db 17287 GCGTGGCCC 17279

RESULT 6
US-09-981-353-183
; Sequence 183, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 183
; LENGTH: 5648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 464689.64
US-09-981-353-183

Query Match          9.9%; Score 409.8; DB 9; Length 5648;
Best Local Similarity 99.5%; Pred. No. 3.1e-99;
Matches 411; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3719 GGTGTGTGTCGGTTCGCGAGCATCCCGCGCCCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3838
Db 1 GGTGTGTGTCGGTTCGCGAGCATCCCGCGCCCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 120
QY 3779 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3838
Db 61 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 120
QY 3839 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3838
Db 121 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 180
QY 3899 GTACCCCGTTCGCGCTAGCTTTAAATTCGCGCTCGGGACCTCCACGACCGCGG 3958
Db 181 GTACCCCGTTCGCGCTAGCTTTAAATTCGCGCTCGGGACCTCCACGACCGCGG 240
QY 3959 GTACCCCGTTCGCGCTAGCTTTAAATTCGCGCTCGGGACCTCCACGACCGCGG 4018
Db 241 GTACCCCGTTCGCGCTAGCTTTAAATTCGCGCTCGGGACCTCCACGACCGCGG 300
QY 4019 CGAAACCGGAGTCTCTCCGCGACCGCAACTCCGCTCCGAGGCTCAGCCCGCTGGAAG 4078
Db 301 CGAAACCGGAGTCTCTCCGCGACCGCAACTCCGCTCCGAGGCTCAGCCCGCTGGAAG 360
QY 4079 TGATCCCGCATCCGAGAGCAAGATGCGCGGCCACTTGTCTGAGGACGATGT 4131
Db 361 TGATCCCGCATCCGAGAGCAAGATGCGCGGCCACTTGTCTGAGGACGATGT 413

RESULT 7
US-09-919-172-70
; Sequence 70, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 70
; LENGTH: 5648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 464689.64
US-09-919-172-70

Query Match          9.9%; Score 409.8; DB 10; Length 5648;
Best Local Similarity 99.5%; Pred. No. 3.1e-99;
Matches 411; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3719 GGTGTGTGTCGGTTCGCGAGCATCCCGCGCCCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3778
Db 1 GGTGTGTGTCGGTTCGCGAGCATCCCGCGCCCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 60
QY 3779 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3838
Db 61 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 120
QY 3839 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3838
Db 121 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 180
QY 3899 GTACCCCGTTCGCGCTAGCTTTAAATTCGCGCTCGGGACCTCCACGACCGCGG 3958
Db 181 GTACCCCGTTCGCGCTAGCTTTAAATTCGCGCTCGGGACCTCCACGACCGCGG 3958
```

Db 181 GTACCCGTTGCGAGCTCTAGCCCTTAAATTCCTCCGGCTCGGGACCTCCACGACCCGGG 240  
QY 3959 CTAGCCGCGACACACAGCTAGCTGCAAGCGCGCGGCTCAGCGGTACCGGCGGCTT 4018  
Db 241 CTAGCCGCGACACACAGCTAGCTGCAAGCGCGCGGCTCAGCGGTACCGGCGGCTT 300  
QY 4019 CGAAACCGAGTCTCTCGGCGACCCCGAACTCCGCTCCGAGAGCTCAGCCCTCGAAAG 4078  
Db 301 CGAAACCGAGTCTCTCGGCGACCCCGAACTCCGCTCCGAGAGCTCAGCCCTCGAAAG 360  
QY 4079 TGTATCCCGGCATCCGAGAGCCAGATGCGCGGCCACTTCTCGAGAGCATGT 4131  
Db 361 TGTATCCCGGCATCCGAGAGCCAGATGCGCGGCCACTTCTCGAGAGCATGT 413  
RESULT 8  
US-09-764-891-7810  
; Sequence 7810, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7810  
; LENGTH: 13216  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-7810

Query Match 8.4%; Score 349; DB 9; Length 13216;  
Best Local Similarity 70.0%; Pred. No. 1,3e-82;  
Matches 516; Conservative 0; Mismatches 210; Indels 11; Gaps 3;  
QY 1441 TTTTCTTTTCTTTTGTGAGAAAGAGTCTTCTTTCACCCCTGGCTGAGGCGAGTGTG 1500  
Db 3625 TCTTTCTTTTCTTTTCTGAGTCTTCTTCTGAGTCTTCTGCGCCAGGCTGAGTGTG 3684  
QY 1501 TGATCTCAGCTCAGTCAACTTTCGCTCCCGGTTCAAGCAATCTCTCTGCTCAGCT 1560  
Db 3685 CGATCTCAGCTCAGTCAACTTTCGCTCCCGGTTCAAGCAATCTCTCTGCTTGTGCT 3744  
QY 1561 CCCAAGTAGCTGGGACTACAGGCACTTCCACCATGCCAGCTGATTTTGTATTTTAG 1620  
Db 3745 CCTGAGTAGCTGGGATTACAGGCGTGCATACCGTGCCTGGCTAATTTTATATTTTAA 3804  
QY 1621 TAGAGATGGGATTTTCACTTTGTTGGCCAGCTGGTCTCAAACTTTTGTGTCATAATG 1680  
Db 3805 TACAGATGGGTTTCCACATGTTGTCAGGCTGGTCTCGAACTCCTGACTTCATGATCCG 3864  
QY 1681 TTGTAACTATTCTTCTTTGCTGAGGTAGGCGCCCGACCAAAATAAATCTTAG 1740  
Db 3865 TCCCCCTCGACTCCCAAGTGTAGGATTACAGGATGAGCCACTGTGCCAGCC---- 3920  
QY 1741 AATCAAAATCAGTGTGTGGTTGACCACTGTCACTTGAGAACCACTGTGACCCAGGCG 1800  
Db 3921 ---GCTAATTTTATTTTGTACAGACAGATTTCTCCATGTTTGCCAGGCTGATCGAA 3977  
QY 1801 CTCAGAGTAGAGGTGATCTCTGCTCGAAGAGAAATAGATGAATAATTTCTCCGGCC 1860  
Db 3978 CTCCTGGGCGCAAGTGAACCACTCTCCCTCGGC---CTCAAAAGTCTGGGATTACAGGC 4034  
QY 1861 AGCGTGTGGTCTATGCTGCTGAATCCAGCACTTTGGGAGCCCAAGCATGTGATCAC 1920  
Db 4035 GGGTGGTGGTCTAGGCTGTAAATCCAGCACTTTGGGAGCCCAAGCATGTGATCAC 4094  
QY 1921 CTGAGGTGAGGAGTTCAAAACAGGCTGGCCAACTGTGAAACCCCGTCTCTACTAAA 1980  
Db 4095 CTGAGGTGAGGAGTTCAAGACCCAGCTGGCCAACTGTGAAACCCCGTCTCTACTAAA 4154

QY 1981 ATACAAAAATTAAGTAAAGTGTGTCGCGCATGCTCTATATCCAGTACTTGGGAGGT 2040  
Db 4155 ATAC-AAAAATTAAGCAGTGTGTCGCGCATGCTCTATATCCAGTACTTGGGAGGT 4213  
QY 2041 GAGGAGGAGAAATTTCTTGAACCCGGGAGGAGAGTGTGAGTGAAGCGAGATCACACCA 2100  
Db 4214 GAGGAGGAGAAATTTCTTGAACCCGGGAGGAGAGTGTGAGTGAAGCGAGATCACACCA 4273  
QY 2101 CTGCACTCCAGCTCGGGGAGAGAGGAGACTTCTCTCAAAAAACAAAAACAAAAA 2160  
Db 4274 CTGCACTCCAGCTCGGGGAGAGAGTGTGAGTGTCTCTCAAAAAACAAAAACAAAAA 4333  
QY 2161 ATTAAGCAAAATTAGACA 2177  
Db 4334 ACAAAACAAACAAACA 4350  
RESULT 9  
US-09-764-891-7809  
; Sequence 7809, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7809  
; LENGTH: 29163  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-7809

Query Match 8.4%; Score 347.4; DB 9; Length 29163;  
Best Local Similarity 69.9%; Pred. No. 5.8e-82;  
Matches 515; Conservative 0; Mismatches 211; Indels 11; Gaps 3;  
QY 1441 TTTTCTTTTCTTTTGTGAGAAAGAGTCTTCTTTCACCCCTGGCTGAGGCGAGTGTG 1500  
Db 3623 TCTTTCTTTTCTTTTCTGAGTCTTCTTCTGAGTCTTCTGCGCCAGGCTGAGTGTG 3682  
QY 1501 TGATCTCAGCTCAGTCAACTTTCGCTCCCGGTTCAAGCAATCTCTCTGCTCAGCT 1560  
Db 3683 CGATCTCAGCTCAGTCAACTTTCGCTCCCGGTTCAAGCAATCTCTCTGCTTGTGCT 3742  
QY 1561 CCCAAGTAGCTGGGACTACAGGCACTTCCACCATGCCAGCTGATTTTGTATTTTAG 1620  
Db 3743 CCTGAGTAGCTGGGATTACAGGCGTGCATACCGTGCCTGGCTAATTTTATATTTTAA 3802  
QY 1621 TAGAGATGGGATTTTCACTTTGTTGGCCAGCTGGTCTCAAACTTTTGTGTCATAATG 1680  
Db 3803 TACAGATGGGTTTCCACATGTTGTCAGGCTGGTCTCGAACTCCTGACTTCATGATCCG 3862  
QY 1681 TTGTAACTATTCTTCTTTGCTGAGGTAGGCGCCCGACCAAAATAAATCTTAG 1740  
Db 3863 TCCCCCTCGACTCCCAAGTGTAGGATTACAGGATGAGCCACTGTGCCAGCC---- 3918  
QY 1741 AATCAAAATCAGTGTGTGGTTGACCACTGTCACTTGAGAACCACTGTGACCCAGGCG 1800  
Db 3919 ---GCTAATTTTATTTTGTACAGACAGATTTCTCCATGTTTGCCAGGCTGATCGAA 3975  
QY 1801 CTCAGAGTAGAGGTGATCTCTGCTCGAAGAGAAATAGATGAATAATTTCTCCGGCC 1860  
Db 3976 CTCCTGGGCGCAAGTGAACCACTCTCCCTCGGC---CTCAAAAGTCTGGGATTACAGGC 4032  
QY 1861 AGCGTGTGGTCTATGCTGCTGAATCCAGCACTTTGGGAGCCCAAGCATGTGATCAC 1920  
Db 4033 GGGTGGTGGTCTACGCTGTAAATCCAGCACTTTGGGAGCCCAAGCATGTGATCAC 4092















GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 14:14:22 ; Search time 5373 seconds

(without alignments)  
12509.077 Million cell updates/sec

Title: US-10-016-725-15

Perfect score: 4150

Sequence: 1 aggaatccatccattaaaat.....tgagttccagcctggtccc 4150

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpi.\*

7: em\_estro.\*

8: em\_hfc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_hfc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: gb\_gss.\*

18: em\_gss\_hum.\*

19: em\_gss\_inv.\*

20: em\_gss\_pln.\*

21: em\_gss\_vrt.\*

22: em\_gss\_fun.\*

23: em\_gss\_mam.\*

24: em\_gss\_mus.\*

25: em\_gss\_other.\*

26: em\_gss\_pro.\*

27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	653.4	15.7	1238	13 BM563325	BM563325 AGENCOURT
2	491.4	11.8	579	13 BI559696	BI559696 603252566
3	482.4	11.6	754	13 BI827092	BI827092 603075763
4	450.2	10.8	471	17 AQ760013	AQ760013 HS_3202_A
5	440.8	10.6	1172	13 BM546583	BM546583 AGENCOURT
6	426.4	10.3	694	9 AL596692	AL596692 DKFZp451P

7	422.4	10.2	469	9	AL701463	AL701463 DKFZp686P
8	421.6	10.2	683	13	BI602109	BI602109 603246216
9	420.8	10.1	560	9	AL708245	AL708245 DKFZp686P
10	407.4	9.8	1012	13	BM465118	BM465118 AGENCOURT
11	406.2	9.8	439	9	AL701687	AL701687 DKFZp686P
12	404.4	9.7	1024	13	BM451306	BM451306 AGENCOURT
13	402.8	9.7	871	14	BQ431921	BQ431921 AGENCOURT
14	402.4	9.7	983	13	BM480208	BM480208 AGENCOURT
15	400.4	9.6	545	9	AL556514	AL556514
16	364.8	8.8	1015	13	BM476059	BM476059 AGENCOURT
17	353.6	8.5	733	13	BI545552	BI545552 603187730
18	329.8	7.9	367	9	AL589552	AL589552 DKFZp451M
19	328.4	7.9	785	9	AU130094	AU130094
20	299.4	7.2	683	9	AU123589	AU123589
21	279.6	6.7	300	9	AU098830	AU098830
22	279.6	6.7	309	9	AU076658	AU076658
23	255.8	6.2	874	13	BM051176	BM051176 603634112
24	255.8	6.2	2971	17	AF101960	AF101960 AF101960
25	249.4	6.0	524	10	AW973992	AW973992 EST385990
26	249.2	6.0	580	14	BQ053726	BQ053726 AGENCOURT
27	248.2	6.0	731	9	AI687343	AI687343 tp86f08.X
28	247.2	6.0	775	14	BM905333	BM905333 AGENCOURT
29	246.8	5.9	598	14	BQ017808	BQ017808 UI-H-ED0-
30	246.2	5.9	508	17	AQ418545	AQ418545 RPCI-11-2
31	245.4	5.9	467	17	AQ388230	AQ388230 RPCI11-14
32	245.4	5.9	785	14	BM721013	BM721013 UI-E-BOO-
33	244.8	5.9	663	17	AG174449	AG174449 Pan t10g1
34	244.6	5.9	750	14	BM678642	BM678642 UI-E-BOO-
35	243.8	5.9	2971	17	AF101960	AF101960 AF101960
36	242.6	5.8	493	10	AW328331	AW328331 ds01g11.y
37	242.4	5.8	532	12	BG830539	BG830539 602767127
38	242.4	5.8	861	12	BG676827	BG676827 602623425
39	242.2	5.8	759	14	BQ436286	BQ436286 AGENCOURT
40	241.8	5.8	1750	11	AF119908	AF119908 Homo sapi
41	241.6	5.8	717	17	AG080871	AG080871 Pan t10g1
42	241.6	5.8	967	12	BE897079	BE897079 601439635
43	241.4	5.8	585	13	BM045436	BM045436 603623552
44	240.8	5.8	535	17	AQ395337	AQ395337 CITR1-E1-
45	240.8	5.8	797	12	BG676969	BG676969 602623602

ALIGNMENTS

RESULT 1  
BM563325  
LOCUS BM563325 1238 bp mRNA linear EST 20-FEB-2002  
DEFINITION AGENCOURT\_6564864 NIH\_MGC\_119 Homo sapiens cDNA clone IMAGE:5743936  
5', mRNA sequence.  
ACCESSION BM563325  
VERSION BM563325.1 GI:18810153  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1238)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12764 row: h column: 17  
High quality sequence stop: 622.  
Location/Qualifiers  
1. .1238

FEATURES  
source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5743936"  
 /clone\_lib="NIH\_MGC\_119"  
 /tissue\_type="medulla"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH\_MGC Library."  
 BASE COUNT 224 a 418 c 370 g 226 t  
 ORIGIN

Query Match 15.7%; Score 653.4; DB 13; Length 1238;  
 Best Local Similarity 98.6%; Pred. No. 1.2e-66;  
 Matches 701; Conservative 0; Mismatches 6; Indels 4; Gaps 4;  
 QY 3266 TGGGTGGCTGCCAGCTCCCTGCCACGCGCTGGGGGACGGGAGACGGGACGAGA 3325  
 Db 20 TGGGTGGCTGCCAGCTCCCTGCC-CGCGCCTGGGGGACGGGAGACGGGACGAGA 78  
 QY 3326 TGTAGTGGTGGCGCCCGCCGAGGGTTACCACTGTTTCTTGAGAAACTTCCCGAGTGC 3385  
 Db 79 TGTAGTGGTGGCGCCCGCCGAGGGTTACCACTGTTTCTTGAGAAACTTCCCGAGTGC 138  
 QY 3386 CCACCCACCGCTTCCTGGTGGCCGAGGCGGCTCGTGGCTAGGCTCGCGCCCCAGC 3445  
 Db 139 CCACCCACCGCTTCCTGGTGGCCGAGGCGGCTCGTGGCTAGGCTCGCGCCCCAGC 198  
 QY 3446 CCACAAACCGGTCCCGAGCCCTTCACAGAGAAAGCTCCGACGGGGATGCCGGCAG 3505  
 Db 199 CCACAAACCGGTCCCGAGCCCTTCACAGAGAAAGCTCCGACGGGGATGCCGGCAG 258  
 QY 3506 AGCCCCAGCGCGGTGGAGAGAAAGCTGAGAGGAGAAACAGAGGGGAGGGGAGCGAG 3565  
 Db 259 AGCCCCAGCGCGGTGGAGAGAAAGCTGAGAGGAGAAACAGAGGGGAGGGGAGCGAG 318  
 QY 3566 GAGCTGGCGCAGAGGAGACAGAGATGGCGCGAGCCATGSCACGGCAGGACGAGGT 3625  
 Db 319 GAGCTGGCGCAGAGGAGACAGAGATGGCGCGAGCCATGSCACGGCAGGACGAGGT 378  
 QY 3626 GGCACCAAAATTCCTTCGGCCAAATGACGAGCGGAGTTTACAGAAGCCTCATTTAGCATTT 3685  
 Db 379 GGCACCAAAATTCCTTCGGCCAAATGACGAGCGGAGTTTACAGAAGCCTCATTTAGCATTT 438  
 QY 3686 CCCACAGGCGAGGCGAGGCGCAGAGCGGGTGGTGTGGTGTGCGGTGTCGGCAGCATCC 3745  
 Db 439 CCCACAGGCGAGGCGAGGCGCAGAGCGGGTGGTGTGGTGTGCGGTGTCGGCAGCATCC 498  
 QY 3746 CCGSGCCCTGCTGCGGTGCCCGAGCCTCGCCTCTGCTCT-CCCGCTCCGCGCTT 3804  
 Db 499 CCGSGCCCTGCTGCGGTGCCCGAGCCTCGCCTCTGCTCTCCCGCTCCGCGCTT 558  
 QY 3805 TACTTCACCGGGACCGCCCGCCAGTCACTTCCTCGCACTTTGCCCCCTGCTTGGCAG 3864  
 Db 559 TACTTCACCGGGACCGCCCGCCAGTCACTTCCTCGCACTTTGCCCCCTGCTTGGCAG 618  
 QY 3865 CGGATAAAGGGGCTGAGAAATACCGGACGGTACCCGTTGCCAGCTTAGCCCTTT 3924  
 Db 619 CGGATAAAGGGGCTGAGAAATACCGGACCGTACCCGTTGCCAGCTTAGCCCTTT 678  
 QY 3925 AATTCCCGCTGCGGAGCTTCACGACAC-GGGGCTAGCGCGGACAC 3973  
 Db 679 TAATTCCCGGCTCGGGAGCTTCCCGACCCGGGGGCTAGCGCGGACAC 729

RESULT 2  
 BI559696

LOCUS  
 DEFINITION 60325256FF1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5294921 5', mRNA linear EST 05-SEP-2001  
 mRNA sequence.  
 ACCESSION BI559696  
 VERSION BI559696  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 579)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM11746 row: c column: 18  
 High quality sequence stop: 576.  
 Location/Qualifiers  
 1..579  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5294921"  
 /clone\_lib="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site: 1: BamHI; Site: 2: SalI; XhoI (gtcag ); Oligo-dT primed using primer 5'-TTTTTATTTTNN-3', normalized to R01 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 103 a 188 c 205 g 82 t 1 others

Query Match 11.8%; Score 491.4; DB 13; Length 579;  
 Best Local Similarity 97.6%; Pred. No. 8.4e-48;  
 Matches 562; Conservative 0; Mismatches 7; Indels 7; Gaps 6;  
 QY 3229 GGAGCGGGGCTTCCCGCGTCCCAAGCTCCAGATCCTGGGGTGGCTGCCACGTCTCCCT 3288  
 Db 4 GGAGCGGGGCTTCCCGCGTCCCAAGCTCCAGATCCTGGGGTGGCTGCCACGTCTCCCT 63  
 QY 3289 GCCACGGCTTGGGGGACGGGAGACGGGACGAGATGTTAGT-GGTGGGCGCCCCCG 3347  
 Db 64 GCCACGGCTTGGGGGACGGGAGACGGGACGAGATGTTAGTGGTGGGCGCCCCCG 123  
 QY 3348 AGGG--TTCACCACTGTTTCTTGAGAACTTCCAGTGCCACCCACCCGTTCTCGTG 3405  
 Db 124 AGGGTTTACNCACTGTTTCTTGAGAACTTCCAGTGCCACCCACCCGTTCTCGTG 183  
 QY 3406 TGCCCGAGGGGCTTCCGTGGGCTAGGCTCGCGCCCCAGCCCCCAACCCGGGTCCCCAGCC 3465  
 Db 184 TGCCCGAGGGGCTTCCGTGGGCTAGGCTCGCGCCCCAGCCCCCAACCCGGGTCCCCAGCC 243  
 QY 3466 CTTCCAGAGAGAAAGCTCCGACGGGATGCCGGGACGAGCCCGAGGGCGGGTGGAA 3525  
 Db 244 CTTCCAGAGAGAAAGCTCCGACGGGATGCCGGGACGAGCCCGAGGGCGGGTGGAA 303  
 QY 3526 GAGAAGCTGAGAGAGAAACAGAGGGAGGG-GGAGCGAGGAGCTGGCGGACAGGGAA 3584  
 Db 304 GAGAAGCTGAGAGAGAAACAGAGGGAGGGCGGAGGAGCTGGCGGACAGGGAA 363

```

QY 3585 CAGCAGATTGGCGGAGCCAAATGGCAA-CGGCAGGAGGAGGTGGCACCACAAATTCCTTCG 3643
Db 364 CAGCAGATTGGCGGAGCCAAATGGCAA-CGGCAGGAGGAGGTGGCACCACAAATTCCTTCG 423
QY 3644 GCCAATGACGAGCGGAGTTTACAGAGCCTCATTAGCATTTCCCGAGAGGC-AGGGGCA 3702
Db 424 GCCAATGACGAGCGGAGTTTACAGAGCCTCATTAGCATTTCCCGAGAGGTAGGGGCA 483
QY 3703 GGGCAGAGCGCGGTTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 3762
Db 484 GGGCAGAGCGCGGTTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 543
QY 3763 TCGCGGAGCGCTCGGCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3798
Db 544 TCGCGGAGCGCTCGG-CTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 578

RESULT 3
BI827092
LOCUS 603075763F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5167695 5',
DEFINITION mRNA sequence.
ACCESSION BI827092
VERSION BI827092.1 GI:15938629
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 754)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cyapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11416 row: n column: 16
High quality sequence stop: 741.
FEATURES
    source
        1..754
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5167695"
                /clone_lib="NIH_MGC_119"
                /tissue_type="medulla"
                /lab_host="DH10B"
                /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: EcoRV (destroyed); RNA source normal medulla from
                anonymous male age 27. Library is oligo-dT primed and
                directionally cloned (EcoRV site is destroyed upon
                cloning). Average insert size 1.3 kb, insert size range
                0.9-3 kb. Library is normalized and enriched for
                full-length clones and was constructed by C. Gruber
                (Invitrogen). Research Genetics tracking code 013. Note:
                this is a NIH_MGC Library."
BASE COUNT 158 a 255 c 208 g 133 t
ORIGIN

Query Match 11.6%; Score 482.4; DB 13; Length 754;
Best Local Similarity 99.4%; Pred. No. 7.5e-47;
Matches 505; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 3613 GGACGAGCAGAGGTGGCACCACAAATTCCTTCGGCCAAATGACGAGCGCGGAGTTACAGAAGC 3672
Db 1 GGACGAGCAGAGGTGGCACCACAAATTCCTTCGGCCAAATGACGAGCGCGGAGTTACAGAAGC 60
QY 3673 CTCATTAGCATTTCCCGACAGAGCGGGGCGACAGCGCGGTTGGTGTGGTGTGGTGTGGT 3732

```

```

Db 61 CTCATTAGCATTTCCCGACAGGCA-GGCGAGGGGACAGAGCGGGTGTGGTGTGGTGTGGT 119
QY 3733 GTCGGCAGCATCCCGCGGCCCTCTGTGGGTGCGCGGAGCCTCGGCTCTGTGTCTCTCTCC 3792
Db 120 GTCGGCAGCATCCCGCGGCCCTCTGTGGGTGCGCGGAGCCTCGGCTCTGTGTCTCTCTCC 179
QY 3733 CCCTCCCGCCCTTACCTCCACGCGGGAGCCGCGCGGCGCAGTCAACTCTCGCACTTGGC 3852
Db 130 CCCTCCCGCCCTTACCTCCACGCGGGAGCCGCGCGGCGCAGTCAACTCTCGCACTTGGC 239
QY 3853 CCTGCTTGGCAGCGATAAAGGGGCTGAGGAATACCGACACGGTCAACCGGTTGGCA 3912
Db 240 CCTGCTTGGCAGCGATAAAGGGGCTGAGGAATACCGACACGGTCAACCGGTTGGCA 299
QY 3913 GCTCTAGCCTTTAAATTCCTCGGCTCGGGGACCTCCACGACCGCGGCTAGCGCGGACAC 3972
Db 330 GCTCTAGCCTTTAAATTCCTCGGCTCGGGGACCTCCACGACCGCGGCTAGCGCGGACAC 359
QY 3973 CAGCTAGGTCAGAGCGCGCGGCTCAGCGGCTACCGCGGCGCTTCAAAACCGCAGTCC 4032
Db 360 CAGCTAGGTCAGAGCGCGCGGCTCAGCGGCTACCGCGGCGCTTCAAAACCGCAGTCC 419
QY 4033 TCCGCGGACCGCGGAACTCC-GCTCGGAGCCTCAGCCCGCTGGAAAGTGATCCCGGCATC 4091
Db 420 TCCGCGGACCGCGGAACTCCCGGCTCAGCGGCTCAGCCCGCTGGAAAGTGATCCCGGCATC 479
QY 4092 CGAGAGCCAAAGATCCCGGCGCCTTGCT 4119
Db 480 GGAGAGCCAAAGATCCCGGCGCCTTGCT 507

RESULT 4
AQ760013
LOCUS HS_3202_A1_B01_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3202 Col=1 Row=C, DNA sequence.
ACCESSION AQ760013
VERSION AQ760013.1 GI:5625489
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 471)
AUTHORS Mahairas,G.C., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3202 row: C column: 1
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 471.
FEATURES
    source
        1..471
            Location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="plate=3202 Col=1 Row=C"
                /clone_lib="CIT Approved Human Genomic Sperm Library D"
                /sex="male"
                /note="Organ: sperm; Vector: pBelBAC11; BAC Clones in

```

FEATURES  
source







```

source      1..439
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="DKFZP686B07148"
            /clone_lib="686 (synonym: hlcc3)"
            /tissue_type="human skeletal muscle"
            /dev_stage="adult"
            /lab_host="DH108"
            /note="vector: pTrioPLEX2; Site_1: SfiIA; Site_2: SfiIB;
                    cdna-collection"
BASE COUNT   69 a    169 c    129 g    72 t
ORIGIN
Query Match          9.8%; Score 406.2; DB 9; Length 439;
Best Local Similarity 99.1%; Pred. No. 6.4e-38;
Matches 419; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 3709 GAGCCCGGGTGTTGGTGTGCTGGCAGCATCCCCGGCGCCCTGTCGCGTGCACG 3768
        |||||
Db 2 GAGCCCGGGTGTTGGTGTGCTGGCAGCATCCCCGGCGCCCTGTCGCGTGCACG 61
        |||||
QY 3769 CGAGCCCTCGGCCTGTGTCCTCTCCGCCCTCCGCCCTTACCCTCCACGCGGGACGCCGCCGCG 3828
        |||||
Db 62 CGAGCCCTCGGCCTGTGTCCTCTCCGCCCTCCGCCCTTACCCTCCACGCGGGACGCCGCCGCG 121

```

```

QY 3829 CCAGTCAACTCTCGACCTTTCGCCCTTGGTGGAGCGGATATAAGAGGGGGCTGAGAAAT 3888
Db 122 CCAGTCAACTCTCTCGACCTTTCGCCCTTGGTGGAGCGGAT-AAAGGGGGCTGAGAAAT 180
QY 3889 ACCGACACGGGTACCGGTGCGACGCTTAGCTTTAAATCCCGGCTCGGGGACTCCA 3948
Db 181 ACCGACACGGGTACCGGTGCGACGCTTAGCTTTAAATCCCGGCTCGGGGACTCCA 240
QY 3949 CGCACCGGGGTAGCGCGACAAACAGCTAGCGTCAAGGGCGCGCGGTCAAGCGCGTAC 4008
Db 241 CGCACCGGGGTAGCGCGACAAACAGCTAGCGTCAAGGGCGCGCGGTCAAGCGCGTAC 300
QY 4009 CGCGGGCTTGAACCGGAGTCTCCCGGGACCCCGAATCCCGTCCGGAGCCTCAGCC 4068
Db 301 CGCGGGCTTGAACCGGAGTCTCCCGGGACCCCGAATCCCGTCCGGAGCCTCAGCC 360
QY 4069 CCTCGGAAGTATCCCGGATCCGAGAGCAAGATGCCGCCACTTGTCTCAGAGAGA 4128
Db 361 CCTCGGAAGTATCCCGGATCCGAGAGCAAGATGCCGCCACTTGTCTCAGAGAGA 420
QY 4129 TGT 4131
Db 421 TAT 423

RESULT 12
BM451306 1024 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6386406 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5493817
DEFINITION 5', mRNA sequence.
ACCESSION BM451306
VERSION BM451306.1 GI:18500346
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1024)
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM2118 row: c column: 02
High quality sequence stop: 564.
FEATURES
source
Location/Qualifiers
1..1024
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5493817"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 219 a 338 c 252 g 214 t 1 others
ORIGIN
Query Match 9.7%; Score 404.4; DB 13; Length 1024;
Best Local Similarity 99.8%; Pred. No. 5.6e-38;
Matches 405; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3726 TGTCGGTGTGCGACGATCCCGCGCGCTGCTGGGTGCGCGAGCCTCGGCCTCTGT 3785
Db 122 TGTCGGTGTGCGACGATCCCGCGCGCTGCTGGGTGCGCGAGCCTCGGCCTCTGT 60
QY 3786 CTTCTCCCTCCCGCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCTCGCA 3845
Db 61 CTTCTCCCTCCCGCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCTCGCA 120
QY 3846 CTTTCCCTCCCTGTTGGCAGCGGATAAAAGGGGGCTGAGAAATACGGACACGGTCACCC 3905
Db 121 CTTTCCCTCCCTGTTGGCAGCGGATAAAAGGGGGCTGAGAAATACGGACACGGTCACCC 180
QY 3906 GTTCCAGCTCTAGCTTTAAATCCCGGCTCGGGGAACTCCACGACCGCGGCTAGGCG 3965
Db 181 GTTCCAGCTCTAGCTTTAAATCCCGGCTCGGGGAACTCCACGACCGCGGCTAGGCG 240
QY 3966 GCACAACAGCTAGCTGCAAGGCGCGGCTCAGCGGTACGCGGGCTTCGAAACC 4025
Db 241 GCACAACAGCTAGCTGCAAGGCGCGGCTCAGCGGTACGCGGGCTTCGAAACC 300
QY 4026 GCAGTCTCCCGCGACCCCGAACTCCGCTCCGAGCCTCAGCCCTCGAAAGTATGCC 4085
Db 301 GCAGTCTCCCGCGACCCCGAACTCCGCTCCGAGCCTCAGCCCTCGAAAGTATGCC 360
QY 4086 GGCATCCGAGAGCAAGATGCCGCCACTTGTCTCAGAGAGATGT 4131
Db 361 GGCATCCGAGAGCAAGATGCCGCCACTTGTCTCAGAGAGATGT 406

RESULT 13
BM431921 871 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7826820 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6153075
DEFINITION 5', mRNA sequence.
ACCESSION BM431921
VERSION BM431921.1 GI:21170997
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 871)
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13491 row: p column: 04
High quality sequence stop: 565.
FEATURES
source
Location/Qualifiers
1..871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6153075"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 183 a 294 c 218 g 176 t
ORIGIN
Query Match 9.7%; Score 402.8; DB 14; Length 871;
Best Local Similarity 99.5%; Pred. No. 9.5e-38;
Matches 404; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 3726 TGTGGTGTGGCAGCATCCCGGGCCCTGCTGCGGTGCGCGAGAGCTCGGCTCTGT 3785
Db 1 TGTGGTGTGGCAGCATCCCGGGCCCTGCTGCGGTGCGCGAGAGCTCGGCTCTGT 60

QY 3786 CTCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCCCGCGGCGAGTCACTCTCCGA 3845
Db 61 CTCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCCCGCGGCGAGTCACTCTCCGA 120

QY 3846 CTTTGCCCTCTGTGGCAGCGGATAAAGAGGGGCTGAGGAAATACCGACACGGTCAACC 3905
Db 121 CTTTGCCCTCTGTGGCAGCGGATAAAGAGGGGCTGAGGAAATACCGACACGGTCAACC 180

QY 3906 GTTGCAGCTCTAGCTTTAAATCCCGGCTCGGGGAGCTCCAGCACCGCGGTAGCGC 3965
Db 181 GTTGCAGCTCTAGCTTTAAATCCCGGCTCGGGGAGCTCCAGCACCGCGGTAGCGC 240

QY 3966 CGACAACAGCTAGGCTCAAGGGCCCGGCTCAGCGCGTACCGCGGGCTTCGAACCC 4025
Db 241 CGACAACAGCTAGGCTCAAGGGCCCGGCTCAGCGCGTACCGCGGGCTTCGAACCC 300

QY 4026 GCAGTCTCTCCGGCGACCCCGAATCCGCTCGGAGCCCTCAGCCCTCGAAAGTGATCCC 4085
Db 301 GCAGTCTCTCCGGCGACCCCGAATCCGCTCGGAGCCCTCAGCCCTCGAAAGTGATCCC 360

QY 4086 GGCATCCGAGAGCCCAAGATGCGCGGCCACTTGTCTCAGGAGCATGT 4131
Db 361 GGCATCCGAGAGCCCAAGATGCGCGGCCACTTGTCTCAGGAGCATGT 406

RESULT 14
BM480208
LOCUS BM480208
DEFINITION AGENCOURT_6468359 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574129
5', mRNA sequence.
ACCESSION BM480208
VERSION BM480208.1 GI:18529250
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaaps-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM2322 row: e column: 10
High quality sequence stop: 659.
FEATURES
source
1. .983
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="IMAGE:5574129"
/tissue_type="NIH_MGC_88"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 207 a 330 c 242 g 204 t
ORIGIN
Query Match 9.7%; Score 402.4; DB 13; Length 983;

QY 3726 TGTGGTGTGGCAGCATCCCGGGCCCTGCTGCGGTGCGCGAGAGCTCGGCTCTGT 3785
Db 1 TGTGGTGTGGCAGCATCCCGGGCCCTGCTGCGGTGCGCGAGAGCTCGGCTCTGT 60

QY 3786 CTCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCCCGCGGCGAGTCACTCTCCGA 3845
Db 61 CTCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCCCGCGGCGAGTCACTCTCCGA 120

QY 3846 CTTTGCCCTCTGTGGCAGCGGATAAAGAGGGGCTGAGGAAATACCGACACGGTCAACC 3905
Db 121 CTTTGCCCTCTGTGGCAGCGGATAAAGAGGGGCTGAGGAAATACCGACACGGTCAACC 180

QY 3906 GTTGCAGCTCTAGCTTTAAATCCCGGCTCGGGGAGCTCCAGCACCGCGGTAGCGC 3965
Db 181 GTTGCAGCTCTAGCTTTAAATCCCGGCTCGGGGAGCTCCAGCACCGCGGTAGCGC 240

QY 3966 CGACAACAGCTAGGCTCAAGGGCCCGGCTCAGCGCGTACCGCGGGCTTCGAACCC 4025
Db 241 CGACAACAGCTAGGCTCAAGGGCCCGGCTCAGCGCGTACCGCGGGCTTCGAACCC 300

QY 4026 GCAGTCTCTCCGGCGACCCCGAATCCGCTCGGAGCCCTCAGCCCTCGAAAGTGATCCC 4085
Db 301 GCAGTCTCTCCGGCGACCCCGAATCCGCTCGGAGCCCTCAGCCCTCGAAAGTGATCCC 360

QY 4086 GGCATCCGAGAGCCCAAGATGCGCGGCCACTTGTCTCAGGAGCATGT 4131
Db 361 GGCATCCGAGAGCCCAAGATGCGCGGCCACTTGTCTCAGGAGCATGT 406

RESULT 15
AL556514
LOCUS AL556514
DEFINITION LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK006Y012 5
prime, mRNA sequence.
ACCESSION AL556514
VERSION AL556514.1 GI:12899258
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 546)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. .546
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="CS0DK006Y012"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="Placenta"
/Note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 97 a 195 c 146 g 95 t 13 others
ORIGIN
Query Match 9.6%; Score 400.4; DB 9; Length 546;

```

```

Best Local Similarity 98.5%; Pred. No. 9.7e-38;
Matches 406; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3720 GTGTGGTGTGGTGTGGCAGCATCCCGGGCCCTGCTGCGGTGCGCGAGAGCTCGGC 3779
Db 6 GCGTCCGGTGGTGTGGCAGCATCCCGGGCCCTGCTGCGGTGCGCGAGAGCTCGGC 65

QY 3780 CTCGTCTCCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCCCGCGGCGAGTCAATC 3839
Db 66 CTCGTCTCCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCCCGCGGCGAGTCAATC 125

QY 3840 CTCGCACATTTGCCCTCTGCTTGCACGCGGATAAAGAGGGGCTGAGGAAATACCGACACGG 3899
Db 126 CTCGCACATTTGCCCTCTGCTTGCACGCGGATAAAGAGGGGCTGAGGAAATACCGACACGG 185

QY 3900 TCACCCGTTGCCAGTCTAGCTTTAAATCCCGGCTCGGGGAGCTCCACGACCGCGGC 3959
Db 186 TCACCCGTTGCCAGTCTAGCTTTAAATCCCGGCTCGGGGAGCTCCACGACCGCGGC 245

QY 3960 TAGCGCCGACACCAAGCTAGGCTCAAGCGCGCGGCTCAGCGGTACCGCGGGCTTC 4019
Db 246 TAGCGCCGACACCAAGCTAGGCTCAAGCGCGCGGCTCAGCGGTACCGCGGGCTTC 305

QY 4020 GAAACCGCAGTCTCTCCGGCAGCCCGAATCCGCTCCGAGAGCTCAGCCCCCTGAAAAGT 4079
Db 306 GAAACCGCAGTCTCTCCGGCAGCCCGAATCCGCTCCGAGAGCTCAGCCCCCTGAAAAGT 365

QY 4080 GATCCCGCATCCGAGAGCCCAAGATGCGCGGCCACTTGTCTCAGGAGCATGT 4131
Db 366 GATCCCGCATCCGAGAGCCCAAGATGCGCGGCCACTTGTCTCAGGAGCATGT 417

RESULT 15
AL556514
LOCUS AL556514
DEFINITION LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK006Y012 5
prime, mRNA sequence.
ACCESSION AL556514
VERSION AL556514.1 GI:12899258
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 546)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. .546
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="CS0DK006Y012"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="Placenta"
/Note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 97 a 195 c 146 g 95 t 13 others
ORIGIN
Query Match 9.6%; Score 400.4; DB 9; Length 546;

```

Best Local Similarity 97.5%; Pred. No. 2.5e-37;  
Matches 398; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 3724 GGTGTCGGTGTGGCAGCATCCCGGCGCCCTGTGTCGGTTCGCCGAGCCTCGGCCCTCT 3783  
|||||  
Db 1 GGTGTCGGTGTGGMAGCATCCCGGCGCCCTGTGTCGGTTCGCCGAGCCTCGGCCCTCT 60  
QY 3784 GTCTCTCCCTCCCTCCCGCCCTTACCTCCACGGGACCGCCCGGCCAGTCAACTCCTCG 3843  
|||||  
Db 61 GTCTCTCTCCCTCCCGCCCTTACCTCCACGGGACCGCCCGGCCAGTCAACTCCTCG 120  
QY 3844 CACTTGTGCCCCCTGTGTCAGCGGATAAAAGGGGCTCAGGAATACCGGACACGGTCAC 3903  
|||||  
Db 121 CACTTGTGCCCCCTGTGTCAGCGGATAAAAGGGGCTCAGGAATACCGGACACGGTCAC 180  
QY 3904 CCGTTGCCAGCTCTAGCCTTTAAATTCCCGGCTCGGGACCTCCACGCACCGCGGCTAGC 3963  
|||||  
Db 181 CCGTTGCCAGCTCTAGCCTTTAAATTCCCGGCTCGGGACCTCCACGCACCGCGGCTAGC 240  
QY 3964 GCCGACACCAAGCTAGCTGCAAGCGCGCGGCTCAGCGGTACCGGGGCTTCGAAA 4023  
|||||  
Db 241 GCCGACACCAAGCTAGCTGCAAGCGCGCGGCTCAGCGGTACCGGGGCTTCGAAA 300  
QY 4024 CCGCAGTCTCTCCGGCGACCCCGAACTCCGCTCCCGAGCCTCAGCCCTCGAAAAGTGATC 4083  
|||||  
Db 301 CCGCAGTCTCTCCGGCGAMCCCGAMCTCCGCTCCGGAGCCTCMGCCCTCGAAAAGTGATC 360  
QY 4084 CCGGCATCCGAGAGCCAGATCGCGGCCACTTGTGTCAGGACGATGT 4131  
|||||  
Db 361 CCGGCMTGGAGAGCCMAGATCGCGGCCACTTGTGTCAGGACGATAT 408

Search completed: May 20, 2003, 19:06:17  
Job time : 5452 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:17:20 ; Search time 45 Seconds  
(without alignments)  
47.378 Million cell updates/sec

Title: US-10-016-725-16

Perfect score: 92  
Sequence: 1 MPAHLIQDVSPFAPW 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

```

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	92	100.0	16	ABB76307	Human stearyl-CoA
2	56	60.9	359	AAV59378	Amino acid sequenc
3	56	60.9	359	ABB44583	Human wound healin
4	49	53.3	285	ABG05270	Novel human diagno
5	49	53.3	285	ABG26777	Novel human diagno
6	47	51.1	334	ABG06715	Novel human diagno
7	47	51.1	334	ABG26818	Novel human diagno
8	45	48.9	1675	AAU75109	Clathrin heavy cha
9	44.5	48.4	61	ABP07968	Human ORFX protein
10	43.5	47.3	518	ABB92812	Herbicidally activ

11	43.5	47.3	632	23	AAE20083	Lactobacillus rham
12	43	46.7	437	23	ABP26786	Streptococcus poly
13	42	45.7	96	23	ABP33456	Human ORF2429 prot
14	42	45.7	184	21	AAV70255	Canine angiotensin
15	42	45.7	184	23	AAO17430	Canine endostatin.
16	42	45.7	230	23	AAO17429	Canine pro-endosta
17	42	45.7	255	21	AAV70405	Class 1 fatty acid
18	42	45.7	423	22	ABG06816	Novel human diagno
19	42	45.7	423	22	ABG29289	Novel human diagno
20	41.5	45.1	388	21	ABP32493	S. lavendulae Mit
21	41	44.6	114	23	ABP09080	Human ORFX protein
22	41	44.6	237	22	ABG05288	Novel human diagno
23	41	44.6	423	22	ABG26779	Novel human diagno
24	41	44.6	1146	23	ABB91737	Herbicidally activ
25	40.5	44.0	51	22	AAH85247	Human immune/haema
26	40.5	44.0	257	21	ABG28917	Arabidopsis thalia
27	40.5	44.0	328	21	ABG28916	Arabidopsis thalia
28	40.5	44.0	431	12	AAK14617	Encoded by (interr
29	40	43.5	81	22	AAAG2815	C glutamicum prote
30	40	43.5	128	22	ABB10756	Human ovarian and/
31	40	43.5	128	22	AAH94131	Human reproductive
32	40	43.5	197	14	AAK37919	Cyn di derived fro
33	40	43.5	197	14	AAK37920	Cyn di derived fro
34	40	43.5	200	14	AAK37913	Cyn di derived fro
35	40	43.5	243	8	AAH70483	Sequence encoded b
36	40	43.5	253	14	AAK37921	Cyn di derived fro
37	40	43.5	255	21	AAV70406	Class 2 fatty acid
38	40	43.5	272	14	AAK37917	Cyn di.18 Cynodo
39	39.5	42.9	72	22	ABB43771	Peptide #11277 enc
40	39.5	42.9	72	22	ABB26714	Protein #8713 enco
41	39.5	42.9	72	22	AAH64733	Human brain expres
42	39.5	42.9	72	22	AAH77503	Human bone marrow
43	39.5	42.9	72	22	AAH21433	Peptide #7867 enco
44	39.5	42.9	72	22	AAH37675	Peptide #11712 enc
45	39.5	42.9	72	23	ABG46528	Human peptide enco

#### ALIGNMENTS

RESULT 1  
ABB76307  
ID ABB76307 standard; Peptide; 16 AA.  
XX AC ABB76307;  
XX AC  
XX DT 22-AUG-2002 (first entry)  
XX DT  
XX DE Human stearyl-CoA desaturase N-terminal peptide..  
XX DE  
XX KW Stearyl-CoA desaturase; SCD; enzyme; human; promoter; virucide;  
KW dermatological; cytostatic; immunosuppressive; antiallergic;  
KW antiarthritic; antiinflammatory; cardiovascular; antianaemic;  
KW gene therapy.  
XX  
XX OS Homo sapiens.  
XX OS  
XX PN WC200236780-A2.  
XX PN  
XX PD 10-MAY-2002.  
XX PD  
XX PF 31-OCT-2001; 2001WO-US45199.  
XX PF  
XX PR 31-OCT-2000; 2000US-244508P.  
XX PR 30-OCT-2001; 2001US-0244508.  
XX PR  
XX PA (JOHJ ) JOHNSON & JOHNSON CONSUMER CO INC.  
XX PA  
XX PI Prouty SM, Zhang L, Steen KS;  
XX PI  
XX DR WP.; 2002-471502/50.  
XX DR N-PSDB; ABL57445.  
XX DR

PT New human stearoyl-CoA desaturase gene promoter, useful for treating a  
 PT skin diseases (e.g. acne, psoriasis and rosacea), tumor diseases,  
 PT leukemias, autoimmune diseases, allergies, arthritis, inflammations, or  
 XX organ rejections

PS Disclosure; Fig 1; 53pp; English.

XX The present sequence is that of the N-terminal region of human  
 CC stearoyl-CoA desaturase (SCD), encoded by the 5' region of the SCD  
 CC coding sequence given in ABL57445. The present invention provides  
 CC a promoter of the SCD gene and functional moieties, fragments and  
 CC variants of it, as well as nucleic acid constructs and vectors that  
 CC contain such sequences, and their uses. The promoter may be used  
 CC for selective transgenic expression in various tissues such as the  
 CC skin for treating a skin disease (e.g. acne, psoriasis and rosacea),  
 CC tumours, leukaemia, autoimmune diseases, allergy, arthritis,  
 CC inflammation, organ rejection, graft versus host reaction, diseases  
 CC of the blood coagulation system, cardiovascular diseases, anaemia,  
 CC infections, or damage to the central nervous system.

XX Sequence 16 AA;

Query Match 100.0%; Score 92; DB 23; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-09;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPAPW 16  
 |||||  
 Db 1 MPAHLQDDVSFPAPW 16

# RESULT 2

AA569378  
 ID AAY69378 standard; Protein; 359 AA.

XX AAY69378;

DT 19-JUN-2000 (first entry)

DE Amino acid sequence of human skin stearoyl-CoA desaturase.

XX Mouse; skin; stearoyl-CoA desaturase; SCD; unsaturated fatty acid;  
 KW skin disorder; skin cancer; hypertrichosis; hirsutism; acne;  
 KW atopic dermatitis; alopecia; gene therapy.

XX Homo sapiens.

XX WO200009754-A2.

XX 24-FEB-2000.

PF 12-AUG-1999; 99WO-US18387.

XX 14-AUG-1998; 98US-0096520.

PR 05-AUG-1999; 99US-0096520.

PA (JOHJ ) JOHNSON & JOHNSON CONSUMER CO INC.

XX Stenn K, Prouty SM, Parimoo S, Zhang L;

XX WPI: 2000-224373/19.

DR N-PSDB; AA261576.

XX Novel nucleic acid useful for diagnosing and treating human skin  
 PT disorder comprises sequences encoding human stearoyl-CoA desaturase

XX Claim 10; Fig 8; 91pp; English.

XX The present sequence represents a human skin stearoyl-CoA desaturase  
 CC (SCD) polypeptide. SCD is responsible for the production of  
 CC unsaturated fatty acids, which are required for energy and lipid  
 CC metabolism, membrane structure and signal transduction. The SCD  
 CC polynucleotides and polypeptides are useful for diagnosing a skin

CC disorder by an abnormal level of SCD expression. The polypeptide is  
 CC useful for determining whether an agent increases or decreases the  
 CC expression level or activity of human SCD in skin cells. Such compounds  
 CC are useful for treating human skin disorders such as skin cancer,  
 CC hypertrichosis and hirsutism which is characterized by an excess of  
 CC SCD activity. The SCD polypeptides and polynucleotides are also useful  
 CC for treating human skin disorders such as acne, atopic dermatitis and  
 CC alopecia. The SCD polynucleotide is also useful in gene therapy.

XX Sequence 359 AA;

Query Match 60.9%; Score 56; DB 21; Length 359;  
 Best Local Similarity 90.9%; Pred. No. 0.27;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11  
 |||||  
 Db 1 MPAHLQDDIS 11

# RESULT 3

AB44583  
 ID AB344583 standard; Protein; 359 AA.

XX AB344583;

DT 25-JAN-2002 (first entry)

DE Human wound healing related polypeptide SEQ ID NO 40.

XX Human; mouse; vulnery; dermatological; skin disorder; wound healing;  
 KW gene therapy.

XX Homo sapiens.

XX CA2325226-A1.

XX 17-MAY-2001.

XX 16-NOV-2000; 2000CA-2325226.

PR 17-NOV-1999; 99DE-1055349.

PR 17-DEC-1999; 99US-0172511.

XX 20-JUN-2000; 2000DE-1030149.

XX (SWIT-) SWITCH BIOTECH AG.

XX Regenbogen J, Wolf E, Goppelt A, Werner S, Halle J;

XX WPI: 2001-433142/47.

XX Use of novel polypeptide or its variant or nucleic acid encoding the  
 PT polypeptide for diagnosing and/or preventing and/or treating skin  
 PT disorders and/or treatment in wound healing or for identifying active  
 PT substances

XX Disclosure; Page 196-198; 265pp; English.

XX The invention relates to the use of a polypeptide (ABB44544-ABB44601,  
 CC ABH44608-ABB44623) or its variant or encoding nucleic acid  
 CC (ABA81990-ABA81995, ABA82016-ABA82032) with vulnery and/or  
 CC dermatological activity for the diagnosis, prevention and treatment of  
 CC skin disorders and treatment in wound healing or for the identification  
 CC of pharmacologically active substances. The nucleic acids are useful in  
 CC gene therapy.

CC Note: The printed sequence listing for this specification was incomplete,  
 CC terminating part way through SEQ ID NO 106. The remaining data was  
 CC obtained from EPO data for an equivalent patent (EP1114862).

XX Sequence 359 AA;

Query Match 60.9%; Score 56; DB 22; Length 359;  
 Best Local Similarity 90.9%; Pred. No. 0.27;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11  
|||||:|  
Db 1 MPAHLQDDIS 11

RESULT 4  
ABG05270  
ID ABG05270 standard; Protein; 285 AA.

XX AC ABG05270;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #5261.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.

XX PN WO200175067-A2.  
XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS69457.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -

XX PS Claim 20; SEQ ID No 35629; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human  
XX CC diagnostic amino acid sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 285 AA;

Query Match 53.3%; Score 49; DB 22; Length 285;  
Best Local Similarity 61.5%; Pred. No. 3.3;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 AHLQDDVSFP 15  
|||||:|  
Db 223 AHLLETLISF 235

RESULT 5  
ABG26777  
ID ABG26777 standard; Protein; 285 AA.

XX AC ABG26777;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #26768.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.

XX PN WO200175067-A2.  
XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS90964.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -

XX PS Claim 20; SEQ ID No 57136; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human  
XX CC diagnostic amino acid sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 285 AA;

Query Match 53.3%; Score 49; DB 22; Length 285;  
Best Local Similarity 61.5%; Pred. No. 3.3;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 AHLQDDVSFP 15  
|||||:|



Db 223 AHLLETLISFPSW 235

## RESULT 6

ABG06715  
ID ABG06715 standard; Protein; 334 AA.

XX AC ABG06715;  
XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6706.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS70902.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
XX PS Claim 20; SEQ ID No 37074; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human  
XX CC diagnostic amino acid sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 334 AA;

Query Match 51.1%; Score 47; DB 22; Length 334;  
Best Local Similarity 61.5%; Pred. No. 8.6;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAW 15  
Db 190 AHLHDLPLISFPSW 202

## RESULT 7

ABG26818  
ID ABG26818 standard; Protein; 334 AA.

XX AC ABG26818;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #26809.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS91005.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
XX PS Claim 20; SEQ ID No 57177; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human  
XX CC diagnostic amino acid sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 334 AA;

Query Match 51.1%; Score 47; DB 22; Length 334;  
Best Local Similarity 61.5%; Pred. No. 8.6;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAW 15  
Db 190 AHLHDLPLISFPSW 202

## RESULT 8

AAU75109  
 ID AAU75109 standard; Protein; 1675 AA.  
 XX  
 AC AAU75109;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Clathrin heavy chain.  
 XX  
 KW MAPKAP-K3; AP-3 delta; APP-695; amyloid A4; Hsp8; heat shock protein 8;  
 KW LI130; NY-REN-58; P38 Alpha; P38 Beta; ERK3; KIAA0934; CDK9;  
 KW cell cycle dependent kinase 9; CLK; C-NAP-1; clathrin heavy chain;  
 KW amphiphysin; PN9109; KIAA1106; neurodegenerative disorder;  
 KW non-insulin dependent diabetes mellitus; NIDDM; Alzheimer's disease;  
 KW inflammatory disease; rheumatoid arthritis; inflammatory bowel disorder;  
 KW atherosclerosis; cardiac hypertrophy; hypoxic brain injury;  
 KW yeast two-hybrid; signal transduction pathway; human;  
 KW mitogen activated protein kinase.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 865..1170  
 FT /note= "this region binds the centrosomal Nek-2  
 FT associated protein 1 (C-NAP1) (see ABK13313),  
 FT the bait protein in a yeast two-hybrid assay,  
 FT producing a complex claimed in claim 1"  
 XX  
 XX WO2001198524-A2.  
 XX  
 XX 27-DEC-2001.  
 XX  
 XX 21-JUN-2001; 2001WO-US19762.  
 XX  
 XX 22-JUN-2000; 2000US-213245P.  
 XX  
 XX (MYRI-) MYRIAD GENETICS INC.  
 XX  
 XX Heichman K, Bartel PL;  
 XX WPI; 2002-122287/16.  
 XX  
 XX New protein complexes comprising protein-protein interactions (e.g.  
 PT MAPKAP-K3/AP-3 delta or C-NAP-1/Clathrin HC), useful for diagnosing  
 PT physiological generative disorders or screening drugs for these  
 PT diseases -  
 XX  
 PS Example 10; Page -; 60pp; English.  
 XX  
 CC The invention describes an isolated protein complex, comprising two  
 CC proteins. The protein complex comprises: protein kinase MAPKAP-K3 and  
 CC AP-3 adaptor complex delta sub-unit; MAPKAP-K3 and amyloid A4  
 CC precursor protein, APP-695; MAPKAP-K3 and heat shock protein (Hsp) 8;  
 CC leucine rich LI130 and NY-REN-58; P38 Alpha and P38 Beta; protein kinase  
 CC ERK3 and KIAA0934 (unknown function); ERK3 and cell cycle dependent  
 CC kinase (CDK) 9; ERK3 and protein kinase CLK; C-NAP-1 and Clathrin heavy  
 CC chain; C-NAP-1 and Amphiphysin; C-NAP-1 and novel protein PN9109 or  
 CC C-NAP-1 and KIAA1106 (unknown function) interactions. The protein  
 CC complexes are useful for diagnosing physiological generative disorders,  
 CC drug screening for agents that modulate the interaction of the proteins  
 CC (thus identify drug targets), and identifying additional proteins in the  
 CC pathway common to the proteins. These physiological disorders include  
 CC non-insulin dependent diabetes mellitus (NIDDM), neurodegenerative  
 CC disorders (e.g. Alzheimer's disease), inflammatory diseases (e.g.  
 CC rheumatoid arthritis and inflammatory bowel disorder) and other human  
 CC disease such as atherosclerosis, cardiac hypertrophy and hypoxic brain  
 CC injury. This sequence represents the clathrin heavy chain protein,  
 CC residues 865-1170 of which binds to the bait protein centrosomal Nek-2  
 CC associated protein 1 (C-NAP1) (see ABK13313) in a yeast two-hybrid assay  
 CC for determining components of signal transduction pathways and forms an  
 CC interaction claimed in claim 1 of the invention.  
 CC Note: This sequence does not appear in the specification but has been  
 CC obtained from a reference given in the invention.

XX SQ Sequence 1675 AA;  
 Query Match 48.9%; Score 45; DB 23; Length 1675;  
 Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 MPAHLQDDVSFPAP 15  
 | | | : | | | : |  
 Db 99 MKAHTMTDDVTFWKW 113  
 RESULT 9  
 ABP07988  
 ID ABP07988 standard; Protein; 61 AA.  
 XX  
 AC ABP07988;  
 XX  
 DT 24-JUN-2002 (first entry)  
 XX  
 DE Human ORFX protein sequence SEQ ID NO:15958.  
 XX  
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2001192523-A2.  
 XX  
 XX 06-DEC-2001.  
 XX  
 XX 29-MAY-2001; 2001WO-US10836.  
 XX  
 XX 30-MAY-2000; 2000US-206132P.  
 XX  
 XX 29-AUG-2000; 2000US-228716P.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Shimkets RA, Leach MD;  
 XX WPI; 2002-106308/14.  
 XX  
 XX N-PSDB; ABN23740.  
 XX  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders -  
 XX  
 PS Disclosure; SEQ ID 15958; 1037pp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,

CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 61 AA;

Query Match 48.4%; Score 44.5; DB 23; Length 61;  
 Best Local Similarity 52.9%; Pred. No. 3.4;  
 Matches 9; Conservative 1; Mismatches 4; Indels 3; Gaps 1;

QY 2 PAHL---LQDDVSPAW 15  
 | | | | | | | | | |  
 Db 25 PRLQKGLSADISPPW 41

RESULT 10

ABB92812  
 ID ABB92812 standard; Protein; 518 AA.

XX AC ABB92812;

XX DT 31-MAY-2002 (first entry)

XX DE Herbicidally active polypeptide SEQ ID NO 2023.

XX KW Herbicidal; plant; agriculture; herbicide.

XX OS Arabidopsis thaliana.

XX PN WO200210210-A2.

XX PD 07-FEB-2002.

XX PF 28-AUG-2001; 2001WO-EP09892.

XX PR 28-AUG-2001; 2001WO-EP09892.

XX PA (FARB ) BAYER AG.

XX PI Tietjen K, Weidler M;

XX PS WPI; 2002-269010/31.

XX CC Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -

XX CC Claim 5; SEQ ID NO 2023; 26lpp + Sequence Listing; English.

XX CC The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.

XX Sequence 518 AA;

Query Match 47.3%; Score 43.5; DB 23; Length 518;  
 Best Local Similarity 38.9%; Pred. No. 56;  
 Matches 7; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 MPABL---LQDDVSPAW 15

: | : : : | | | |

Db 171 IFGHIPGKVEDVGFPW 188

RESULT 11

AAE20083  
 ID AAE20083 standard; Protein; 632 AA.

XX AC AAE20083;

XX DT 18-JUN-2002 (first entry)

XX DE Lactobacillus rhamnosus strain HN001 peptidase pepO.

XX KW Enzyme; flavour; aroma; texture; nutritional; dairy manufacture; therapy;  
 KW fermentation process; anti-infection; rotavirus infection; heart disease;  
 KW infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder;  
 KW anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori;  
 KW antihypertensive effect; urogenital infection; hepatic encephalopathy;  
 KW bowel syndrome; endocarditis; transgenic microbe; peptidase; pepO.

XX OS Lactobacillus rhamnosus HN001.

XX PN WO200212506-A1.

XX PD 14-FEB-2002.

XX PF 08-AUG-2001; 2001WO-NZ00160.

XX PR 08-AUG-2000; 2000US-0634238.

XX PR 28-NOV-2000; 2000US-0724623.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

XX PI (VIAL-) VIALACTIA BIOSCIENCE NZ LTD.

XX PI Glenn M, Havukkala IJ, Bloksberg IN, Lubbers WM, Dekker J;

XX PI Christensson AC, Holland R, O'coole PW, Reid JR, Coolbear T;

XX DR WPI; 2002-241760/29.

XX DR N-PSDB; AAD31855.

XX PT New polynucleotides and polypeptides from Lactobacillus rhamnosus,

XX PT useful in e.g. improving the flavor, aroma, texture and health-related

XX PT benefits of milk-derived products, or in increasing properties of

XX PT microbes -

XX PS Claim 11; Fig 81; 257pp; English.

XX CC The present invention relates to a new isolated polynucleotide comprising  
 CC a sequence present in Lactobacillus rhamnosus strain HN001 and encoding a  
 CC polypeptide capable of modifying the flavour, aroma, texture, nutritional  
 CC and health benefits of milk-derived products, and/or survivability of  
 CC microbes in dairy manufacturing processes. The polynucleotides are useful  
 CC for improving the properties of microbes used in the manufacture of milk-  
 CC derived products such as cheeses, yogurt, fermented milk products, sour  
 CC milks and buttermilk; in modifying the flavour, aroma, texture and health  
 CC -related benefits of milk-derived products and in increasing the survival  
 CC of microbes during industrial fermentation processes. The bacteria may be  
 CC used to increase resistance to enteric pathogens and anti-infection  
 CC activity, including treatment of rotavirus infection and infantile  
 CC diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis;  
 CC liver cancer reduction; reduction of small bowel bacterial overgrowth;  
 CC immune system modulation and treatment of autoimmune disorders and  
 CC allergies; treatment of allergic responses to foods; reduction of blood  
 CC lipids and prevention of heart disease; antihypertensive effect;  
 CC prevention and treatment of urogenital infections, Helicobacter pylori,  
 CC or hepatic encephalopathy; treatment of inflammatory bowel disorder and  
 CC irritable bowel syndrome; modulation of endocarditis; and for improved  
 CC protein and carbohydrate utilization and conversion. The transgenic  
 CC microbial population can be administered to a mammal as an anti-  
 CC carcinogenic agent. The present sequence is Lactobacillus rhamnosus  
 CC peptidase pepO.

XX Sequence 632 AA;

Query Match 47.3%; Score 43.5; DB 23; Length 632;  
 Best Local Similarity 42.9%; Pred. No. 70;  
 Matches 9; Conservative 4; Mismatches 1; Indels 7; Gaps 1;

QY 1 MPAHL-----LQDDVSEPA 14  
 ||||| |:::|||||  
 Db 430 MPGLVNASYDPLKNDITFEA 450

RESULT 12  
 ABP26786  
 ID ABP26786 standard; Protein; 437 AA.  
 XX AC ABP26786;  
 XX DT 02-JUL-2002 (first entry)  
 XX DE Streptococcus polypeptide SEQ ID NO 2748.  
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX OS Streptococcus pyogenes.  
 XX PN WO200234771-A2.  
 XX PD 02-MAY-2002.  
 XX PF 29-OCT-2001; 2001WO-GB04789.  
 XX PR 27-OCT-2000; 2000GB-0026333.  
 XX PR 24-NOV-2000; 2000GB-0028727.  
 XX PR 07-MAR-2001; 2001GB-0005640.  
 XX PA (CHIR-) CHIRON SPA.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX WPI; 2002-352536/38.  
 XX N-PSDB; ABN67417.  
 XX DR New Streptococcus protein for the treatment or prevention of infection  
 XX DR or disease caused by Streptococcus bacteria, such as meningitis, and  
 XX DR for detecting a compound that binds to the protein -  
 XX PS Claim 1; Page 3426; 4525pp; English.  
 XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 XX SQ Sequence 437 AA;

Query Match 46.78%; Score 43; DB 23; Length 437;  
 Best Local Similarity 54.5%; Pred. No. 56;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LQDDVSEPAW 15  
 ::::|::|  
 Db 117 IIQDDISESPW 127

RESULT 13  
 ABP33455  
 ID ABP33455 standard; Protein; 96 AA.  
 XX AC ABP33455;  
 XX DT 08-JUL-2002 (first entry)  
 XX DE Human ORF2429 protein, SEQ ID NO:4858.

XX KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;  
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.

XX OS Homo sapiens.

XX WO200190366-A2.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-US17076.

XX 24-MAY-2000; 2000US-206690P.

XX (CJRA-) CURAGEN CORP.

XX Leach MD, Shimkets RA;

XX WPI; 2002-106200/14.

XX N-PSDB; ABN77482.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 XX preventing and treating cardiovascular disease, neurodegenerative,  
 XX hyperproliferative disorders and disorders related to organ  
 XX transplantation -

XX Claim 10; Page 1468; 2508pp; English.

XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 XX designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 XX ABN79587 represent cDNAs encoding them. The invention also encompasses  
 XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 XX referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 XX the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 XX polynucleotides, the recombinant production of ORFX proteins, antibodies  
 XX specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 XX polypeptides, methods of screening for modulators of ORFX expression or  
 XX activity, and methods of screening individuals for a predisposition to an  
 XX ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 XX range of biological activities, such as cytokine, cell proliferation,  
 XX cell differentiation, immune modulation, haematopoiesis regulation,  
 XX tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 XX chemokinetic activity, haemostatic activity, thrombolytic activity,  
 XX receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 XX and antiinfective activity, and may also be involved in the determination  
 XX of bodily characteristics, fertility and behaviour. ORFX proteins,  
 XX nucleic acids and antibodies may be used in the treatment of cancers,

CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.

XX Sequence 96 AA;

SQ Query Match 45.7%; Score 42; DB 23; Length 96;

Best Local Similarity 50.0%; Pred. No. 15; Indels 0; Gaps 0;  
 Matches 8; Conservative 1; Mismatches 7;

QY 1 MPAHLQDDVSEFPWP 16  
 I I I I I  
 Db 60 METHLLVSRVTMAAP 75

RESULT 14

AAAY70265  
 ID AAY70265 standard; Protein; 184 AA.

AC AAY70265;

XX 06-JUN-2000 (first entry)

DE Canine angiogenesis inhibitor, endostatin.

XX Canine; immunoglobulin Fc fragment; endostatin; immunofusin;  
 KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;  
 KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;  
 KW vasotropic; vulnerable; treatment; antiarteriosclerosis; tumour;  
 KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;  
 KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
 KW wound granulation; keloid scar; gene therapy.

XX Canis familiaris.

XX WO200011033-A2.

XX 02-MAR-2000.

XX 25-AUG-1999; 99WO-US19329.

XX 25-AUG-1998; 98US-0097883.

XX (LEXI-) LEXINGEN PHARM CORP.

XX Lo K, Li Y, Gillies SD;

XX WPI; 2000-237616/20.

XX N-PSDB; AAZ51309.

XX Novel fusion protein of angiostatin or endostatin and an immunoglobulin  
 PT FC region, useful for treating conditions mediated by angiogenesis,  
 PT such as rheumatoid arthritis, tumors and macular degeneration -

XX Example 8; Pages 59-60; 68pp; English.

XX The patent discloses a DNA molecule encoding a fusion protein comprising  
 CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis  
 CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment  
 CC having angiostatin activity, a collagen XVIII fragment having endostatin

CC activity, or combinations of them. The fusion protein (immunofusin) is  
 CC used to inhibit angiogenesis and to treat diseases or conditions mediated  
 CC by angiogenesis. Conditions that may be treated include solid tumours,  
 CC blood born tumours, tumour metastasis, benign tumours including  
 CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic  
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases  
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular  
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental  
 CC fibroplasia, rubecosis and Osler-Webber syndrome; myocardial angiogenesis,  
 CC plaque neovascularisation, telangiectasia, haemophilic joints,  
 CC angiofibroma, wound granulation, and excessive or abnormal stimulation of  
 CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and  
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used  
 CC in gene therapy. The present sequence is a canine  
 CC endostatin used in the construction of immunofusin containing canine  
 CC immunoglobulin Fc fragment.

XX Sequence 184 AA;

SQ Query Match 45.7%; Score 42; DB 21; Length 184;

Best Local Similarity 46.7%; Pred. No. 32;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSEFPWP 15

Db 59 VPVNLDRDEVLPSPW 83

RESULT 15

AAO17430

ID AAO17430 standard; Protein; 184 AA.

XX AAO17430;

XX 19-JUL-2002 (first entry)

XX Canine endostatin.

XX Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;  
 KW psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;  
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;  
 KW rubecosis; Osler-Webber syndrome; myocardial angiogenesis;  
 KW plaque neovascularisation; telangiectasia; haemophilic joints;  
 KW angiofibroma; wound granulation; coronary collateral;  
 KW cerebral collateral; arteriovenous malformation;  
 KW ischaemic limb angiogenesis; diabetic neovascularisation; fracture;  
 KW cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;  
 KW gynaecological.

XX Canis familiaris.

XX EP1191036-A2.

XX 27-MAR-2002.

XX 24-AUG-2001; 2001EP-0307224.

XX 25-AUG-2000; 2000US-227924P.

XX (PFIZ ) PFIZER PROD INC.

XX Sheppard MG, Tong X;

XX WPI; 2002-354068/39.

XX N-PSDB; AAL46063.

XX An isolated nucleic acid molecule for the treatment of  
 PT angiogenesis-related disorder, such as cancers or diabetic retinopathy,  
 PT encodes an endostatin protein -

XX Claim 14; Fig 5; 56pp; English.

XX The present invention provides the protein and coding sequences of canine

CC pro-endostatin and endostatin. The sequences can be used in the treatment  
 CC and diagnosis of angiogenesis related disorders, including cancer,  
 CC rheumatoid arthritis, psoriasis, retinopathy, macular degeneration,  
 CC corneal graft rejection, neovascular glaucoma, retrolental fibroplasia,  
 CC rubecosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque  
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,  
 CC wound granulation, coronary collaterals, cerebral collaterals,  
 CC arteriovenous malformations, ischaemic limb angiogenesis, diabetic  
 CC neovascularisation, and fractures. The present sequence is the canine  
 CC pro-endostatin protein sequence.

XX  
 SQ Sequence 184 AA;

Query Match 45.7%; Score 42; DB 23; Length 184;  
 Best Local Similarity 46.7%; Pred. No. 32;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPAW 15  
 :| :|:| :|:  
 Db 69 VPVYLRDEVLPFW 83

Search completed: May 14, 2003, 15:24:17  
 Job time : 57 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:23:29 ; Search time 14 Seconds  
(without alignments)  
33.626 Million cell updates/sec

Title: US-10-016-725-16  
Perfect score: 92  
Sequence: 1 MPAHLLQDDVSFPAPW 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Issued Patents.AA.\*
- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/iaa/PCPUS\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	43.5	68	4	US-08-441-507-9
2	40	43.5	68	4	US-07-969-875A-9
3	40	43.5	86	4	US-08-441-507-8
4	40	43.5	86	4	US-07-969-875A-8
5	40	43.5	197	4	US-08-441-507-5
6	40	43.5	197	4	US-07-969-875A-5
7	40	43.5	200	4	US-08-441-507-4
8	40	43.5	200	4	US-07-969-875A-4
9	40	43.5	246	4	US-08-441-507-21
10	40	43.5	246	4	US-07-969-875A-21
11	40	43.5	272	4	US-08-441-507-15
12	40	43.5	272	4	US-07-969-875A-15
13	39	42.4	356	1	US-08-552-142A-15
14	39	42.4	356	1	US-08-910-973-15
15	39	42.4	356	3	US-09-234-332-7
16	39	42.4	356	3	US-09-234-332-8
17	39	42.4	356	4	US-09-499-227-15
18	39	42.4	357	1	US-08-552-142A-2
19	39	42.4	357	1	US-08-910-973-2
20	39	42.4	357	4	US-09-499-227-2
21	39	42.4	357	5	PCI-US95-05741-2
22	39	42.4	509	4	US-08-857-076-46
23	39	42.4	510	4	US-08-857-076-45
24	39	42.4	635	4	US-08-857-076-101
25	38	41.3	38	1	US-08-176-500-49
26	38	41.3	38	1	US-08-471-052A-49
27	38	41.3	38	1	US-08-189-331-49

28	38	41.3	38	2	US-08-471-939-49	Sequence 49, Appl
29	38	41.3	38	2	US-08-471-800-49	Sequence 49, Appl
30	38	41.3	38	2	US-08-471-068-49	Sequence 49, Appl
31	38	41.3	178	4	US-09-315-689-5	Sequence 5, Appl
32	38	41.3	182	4	US-09-561-500-14	Sequence 14, Appl
33	38	41.3	182	4	US-09-561-108-14	Sequence 14, Appl
34	38	41.3	182	4	US-09-315-689-3	Sequence 3, Appl
35	38	41.3	182	4	US-09-561-526-14	Sequence 14, Appl
36	38	41.3	183	4	US-09-206-059-2	Sequence 2, Appl
37	38	41.3	662	4	US-09-061-768A-25	Sequence 25, Appl
38	37	40.2	47	4	US-09-149-476-359	Sequence 359, Appl
39	37	40.2	216	4	US-09-077-955-31	Sequence 31, Appl
40	37	40.2	256	4	US-09-077-955-30	Sequence 30, Appl
41	37	40.2	294	4	US-09-077-955-29	Sequence 29, Appl
42	37	40.2	338	4	US-09-077-955-28	Sequence 28, Appl
43	37	40.2	358	2	US-08-558-823-19	Sequence 19, Appl
44	37	40.2	368	2	US-08-869-137-2	Sequence 2, Appl
45	37	40.2	390	4	US-09-077-955-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1  
US-08-441-507-9  
; Sequence 9, Application US/08441507  
; Patent No. 6214358  
; GENERAL INFORMATION:  
; APPLICANT: Singh, Mohan Bir;  
; APPLICANT: Smith, Penelope; and  
; APPLICANT: Knox, Robert Bruce  
; TITLE OF INVENTION: Protein Allergens of the Species Cynodon  
; TITLE OF INVENTION: Dactylon  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/441,507  
; FILING DATE: 15-May-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/969,875  
; FILING DATE: 30-October-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mandragouras, Amy E.  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: IMI-049DV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 742-4214  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 68 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
; US-08-441-507-9  
Query Match 43.5%; Score 40; DB 4; Length 68;  
Best Local Similarity 61.5%; Pred. No. 4.9;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSEPAW 15  
Db 42 AHLVQDDVIPANW 54

RESULT 2  
US-07-969-875A-9  
; Sequence 9, Application US/07969875A  
; Patent No. 6441157  
; GENERAL INFORMATION:  
; APPLICANT: Singh, Mohan Bir;  
; APPLICANT: Smith, Penelope; and  
; APPLICANT: Knox, Robert Bruce  
; TITLE OF INVENTION: Protein Allergens of the Species  
; TITLE OF INVENTION: Cynodon Dactylon  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunologic Pharmaceutical Corporation  
; STREET: 610 Lincoln Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/969,875A  
; FILING DATE: 30-October-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Vanstone, Darlene A.;  
; REGISTRATION NUMBER: 35,279  
; REFERENCE/DOCKET NUMBER: 041.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6010  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 68 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
US-07-969-875A-9

Query Match 43.5%; Score 40; DB 4; Length 68;  
Best Local Similarity 61.5%; Pred. No. 4.9;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSEPAW 15  
Db 42 AHLVQDDVIPANW 54

RESULT 3  
US-08-441-507-8  
; Sequence 8, Application US/08441507  
; Patent No. 6214358  
; GENERAL INFORMATION:  
; APPLICANT: Singh, Mohan Bir;  
; APPLICANT: Smith, Penelope; and  
; APPLICANT: Knox, Robert Bruce  
; TITLE OF INVENTION: Protein Allergens of the Species  
; TITLE OF INVENTION: Dactylon  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts

COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441,507  
FILING DATE: 15-May-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/969,875  
FILING DATE: 30-October-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-049DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 86 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-441-507-8

Query Match 43.5%; Score 40; DB 4; Length 86;  
Best Local Similarity 61.5%; Pred. No. 6.4;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSEPAW 15  
Db 50 AHLVQDDVIPANW 72

RESULT 4  
US-07-969-875A-8  
; Sequence 8, Application US/07969875A  
; Patent No. 6441157  
; GENERAL INFORMATION:  
; APPLICANT: Singh, Mohan Bir;  
; APPLICANT: Smith, Penelope; and  
; APPLICANT: Knox, Robert Bruce  
; TITLE OF INVENTION: Protein Allergens of the Species  
; TITLE OF INVENTION: Cynodon Dactylon  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunologic Pharmaceutical Corporation  
; STREET: 610 Lincoln Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/969,875A  
; FILING DATE: 30-October-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Vanstone, Darlene A.;  
; REGISTRATION NUMBER: 35,279  
; REFERENCE/DOCKET NUMBER: 041.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 466-6000  
; TELEFAX: (617) 466-6010



;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 86 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
US-07-969-875A-8

Query Match 43.5%; Score 40; DB 4; Length 86;  
Best Local Similarity 61.5%; Pred. No. 6.4;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAP 15  
| | | | |  
Db 60 AHLVQDDVIPANW 72

RESULT 5  
US-08-441-507-5  
; Sequence 5, Application US/08441507  
; Patent No. 6214358  
; GENERAL INFORMATION:  
; APPLICANT: Singh, Mohan Bir;  
; APPLICANT: Smith, Penelope; and  
; APPLICANT: Knox, Robert Bruce  
; TITLE OF INVENTION: Protein Allergens of the Species Cynodon  
; TITLE OF INVENTION: Dactylon  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: ASCII text  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/441,507  
;; FILING DATE: 15-May-1995  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/969,875  
;; FILING DATE: 30-October-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mandragouras, Amy E.  
;; REGISTRATION NUMBER: 36,207  
;; REFERENCE/DOCKET NUMBER: IMI-049DV  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 227-7400  
;; TELEFAX: (617) 742-4214

;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 197 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
US-08-441-507-5

Query Match 43.5%; Score 40; DB 4; Length 197;  
Best Local Similarity 61.5%; Pred. No. 17;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAP 15  
| | | | |  
Db 173 AHLVQDDVIPANW 185

RESULT 6  
US-07-969-875A-5  
; Sequence 5, Application US/07969875A  
; Patent No. 6441157  
; GENERAL INFORMATION:  
; APPLICANT: Singh, Mohan Bir;  
; APPLICANT: Smith, Penelope; and  
; APPLICANT: Knox, Robert Bruce  
; TITLE OF INVENTION: Protein Allergens of the Species  
; TITLE OF INVENTION: Cynodon Dactylon  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ImmuLogic Pharmaceutical Corporation  
; STREET: 610 Lincoln Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02154

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: ASCII text  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/969,875A  
;; FILING DATE: 30-October-1992  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Vanstone, Darlene A..  
;; REGISTRATION NUMBER: 35,279  
;; REFERENCE/DOCKET NUMBER: 041.1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 466-6000  
;; TELEFAX: (617) 466-6010

;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 197 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: Peptide  
;; FRAGMENT TYPE: internal  
US-07-969-875A-5

Query Match 43.5%; Score 40; DB 4; Length 197;  
Best Local Similarity 61.5%; Pred. No. 17;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAP 15  
| | | | |  
Db 173 AHLVQDDVIPANW 185

RESULT 7  
US-08-441-507-4  
; Sequence 4, Application US/08441507  
; Patent No. 6214358  
; GENERAL INFORMATION:  
; APPLICANT: Singh, Mohan Bir;  
; APPLICANT: Smith, Penelope; and  
; APPLICANT: Knox, Robert Bruce  
; TITLE OF INVENTION: Protein Allergens of the Species Cynodon  
; TITLE OF INVENTION: Dactylon  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441,507  
FILING DATE: 15-May-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/969,875  
FILING DATE: 30-October-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-049DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 200 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-441-507-4

Query Match 43.5%; Score 40; DB 4; Length 200;  
Best Local Similarity 61.5%; Pred. No. 18;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSFPANW 15  
| | | : | | | | |  
Db 174 AHLVQDDVIPANW 186

## RESULT 8

US-07-969-875A-4  
Sequence 4, Application US/07969875A  
Patent No. 6441157

GENERAL INFORMATION:  
APPLICANT: Singh, Mohan Bir;  
APPLICANT: Smith, Penelope; and  
APPLICANT: Knox, Robert Bruce  
TITLE OF INVENTION: Protein Allergens of the Species  
TITLE OF INVENTION: Cynodon Dactylon  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation  
STREET: 610 Lincoln Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02154

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/969,875A  
FILING DATE: 30-October-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Vanstone, Darlene A.  
REGISTRATION NUMBER: 35,279  
REFERENCE/DOCKET NUMBER: 041.1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 466-6000  
TELEFAX: (617) 466-6010  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 200 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-07-969-875A-4

Query Match 43.5%; Score 40; DB 4; Length 200;  
Best Local Similarity 61.5%; Pred. No. 18;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSFPANW 15  
| | | : | | | | |  
Db 174 AHLVQDDVIPANW 186

## RESULT 9

US-08-441-507-21  
Sequence 21, Application US/08441507  
Patent No. 6214358

GENERAL INFORMATION:  
APPLICANT: Singh, Mohan Bir;  
APPLICANT: Smith, Penelope; and  
APPLICANT: Knox, Robert Bruce  
TITLE OF INVENTION: Protein Allergens of the Species Cynodon  
TITLE OF INVENTION: Dactylon  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441,507  
FILING DATE: 15-May-1995  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/969,875  
FILING DATE: 30-October-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: IMI-049DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-441-507-21

Query Match 43.5%; Score 40; DB 4; Length 246;  
Best Local Similarity 61.5%; Pred. No. 22;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSFPANW 15  
| | | : | | | | |  
Db 220 AHLVQDDVIPANW 232

## RESULT 10

US-07-969-875A-21  
Sequence 21, Application US/07969875A  
Patent No. 6441157

GENERAL INFORMATION:  
APPLICANT: Singh, Mohan Bir;  
APPLICANT: Smith, Penelope; and  
APPLICANT: Knox, Robert Bruce

```
; TITLE OF INVENTION: Protein Allergens of the Species
; NUMBER OF INVENTION: Cynodon Dactylon
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunologic Pharmaceutical Corporation
; STREET: 610 Lincoln Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/969,875A
; FILING DATE: 30-October-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vanstone, Darlene A.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 041.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 466-6000
; TELEFAX: (617) 466-6010
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-969-875A-21

Query Match 43.5%; Score 40; DB 4; Length 246;
Best Local Similarity 61.5%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 AHLQDDVSFPW 15
Db 220 AHLVDDVIPANW 232
|||||
|

RESULT 11
US-08-441-507-15
; Sequence 15, Application US/08441507
; Patent No. 6214358
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species Cynodon
; TITLE OF INVENTION: Dactylon
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,507
; FILING DATE: 15-May-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,875
; FILING DATE: 30-October-1992
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 35,207
; REFERENCE/DOCKET NUMBER: IMI-049DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5-8
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15-16
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 42
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 71-72
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; US-08-441-507-15

Query Match 43.5%; Score 40; DB 4; Length 272;
Best Local Similarity 61.5%; Pred. No. 25;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 AHLQDDVSFPW 15
Db 246 AHLVDDVIPANW 258
|||||
|

RESULT 12
US-07-969-875A-15
; Sequence 15, Application US/07969875A
; Patent No. 6441157
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species
; TITLE OF INVENTION: Cynodon Dactylon
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunologic Pharmaceutical Corporation
; STREET: 610 Lincoln Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/969,875A
; FILING DATE: 30-October-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```

; NAME: Vanstone, Darlene A.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 041.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 466-6000
; TELEFAX: (617) 466-6010
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
; FRAGMENT TYPE: Internal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5-8
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15-16
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 42
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 71-72
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; US-07-969-875A-15

```

```

Query Match 43.5%; Score 40; DB 4; Length 272;
Best Local Similarity 61.5%; Pred. NO. 25;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 3 AHLQDDVSFPAP 15
Db 246 AHLQDDVIPANW 258

```

```

RESULT 13
US-08-552-142A-15
; Sequence 15, Application US/08552142A
; Patent No. 5695995
; GENERAL INFORMATION:
; APPLICANT: Weintraub, Harold M.
; APPLICANT: Lee, Jacqueline E.
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Hollenberg, Stanley M.
; TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Genes
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/552,142A
; FILING DATE: 02-NOV-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Broderick, Thomas F.
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: FPCR-1-8933
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-225-0709
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-552-142A-15

```

```

Query Match 42.4%; Score 39; DB 1; Length 356;
Best Local Similarity 50.0%; Pred. NO. 52;
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 MPAHLQDDVSFPAP 16
Db 203 MPPLPTASAFVHP 218

```

```

RESULT 14
US-08-910-973-15
; Sequence 15, Application US/08910973
; Patent No. 5795723
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoderm
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,973
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FPCR-1-10958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 15:

```

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 356 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-08-910-973-15

Query Match 42.4%; Score 39; DB 1; Length 356;  
Best Local Similarity 50.0%; Pred. No. 52;  
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPWP 16  
Db 203 MPPLPTASFPVHP 218

## RESULT 15

US-09-234-332-7  
; Sequence 7, Application US/09234332A  
; Patent No. 6087168  
; GENERAL INFORMATION:  
; APPLICANT: Cedars-Sinai Medical Center  
; APPLICANT: Michel F. Levesque, M.D.  
; APPLICANT: Thomas Neuman, Ph.D.  
; TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO  
; FILE REFERENCE: P07 41494  
; CURRENT APPLICATION NUMBER: US/09/234.332A  
; CURRENT FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Neuro D1 protein; Genbank Accession D82347  
US-09-234-332-7

Query Match 42.4%; Score 39; DB 3; Length 356;  
Best Local Similarity 50.0%; Pred. No. 52;  
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPWP 16  
Db 203 MPPLPTASFPVHP 218

Search completed: May 14, 2003, 15:31:09  
Job time : 15 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:20:36 ; Search time 17 Seconds  
(without alignments)  
86.612 Million cell updates/sec

Title: US-10-016-725-16  
Perfect score: 92  
Sequence: 1 MPAHLQDDVSFPAMP 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues 349150  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published\_Applications\_AA:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	92	100.0	16	12	US-10-016-725-16
2	56	60.9	359	9	US-10-060-036-179
3	56	60.9	359	12	US-10-029-654-12
4	44	47.8	71	9	US-10-001-857-179
5	43.5	47.3	632	9	US-09-971-536-42
6	42	45.7	184	9	US-10-131-241-49
7	42	45.7	255	9	US-10-224-446-28
8	41	44.6	737	9	US-10-228-931-4
9	41	44.6	737	10	US-09-771-161A-195
10	40	43.5	81	9	US-09-738-626-6569
11	40	43.5	128	9	US-09-764-891-2789
12	40	43.5	255	9	US-10-224-446-30
13	39.5	42.9	72	10	US-09-864-761-42012
14	39	42.4	74	9	US-09-910-009A-56
15	39	42.4	74	9	US-09-910-009A-74
16	39	42.4	74	9	US-09-910-009A-138
17	39	42.4	74	9	US-09-910-009A-144
18	39	42.4	74	9	US-09-910-009A-147
19	39	42.4	75	9	US-09-910-009A-201

20	39	42.4	75	10	US-09-764-877-1086	Sequence 1086, Ap
21	39	42.4	76	9	US-09-910-009A-265	Sequence 265, App
22	39	42.4	78	9	US-09-910-009A-14	Sequence 14, Appl
23	39	42.4	78	9	US-09-910-009A-178	Sequence 178, App
24	39	42.4	102	9	US-09-925-299-877	Sequence 877, App
25	39	42.4	102	10	US-09-925-299-877	Sequence 877, App
26	39	42.4	155	9	US-10-121-988-181	Sequence 181, App
27	39	42.4	396	10	US-09-864-761-43232	Sequence 43232, A
28	39	42.4	509	10	US-09-205-658-46	Sequence 46, Appl
29	39	42.4	509	10	US-09-844-353A-46	Sequence 46, Appl
30	39	42.4	510	10	US-09-205-658-45	Sequence 45, Appl
31	39	42.4	510	10	US-09-844-353A-45	Sequence 45, Appl
32	39	42.4	635	10	US-09-844-353A-101	Sequence 101, App
33	39	42.4	704	9	US-10-121-988-94	Sequence 94, Appl
34	39	42.4	858	9	US-10-121-988-92	Sequence 92, Appl
35	39	42.4	1293	10	US-09-815-242-10079	Sequence 10079, A
36	39	42.4	1294	10	US-09-815-242-13724	Sequence 13724, A
37	38	41.3	63	10	US-09-822-540A-1	Sequence 1, Appl
38	38	41.3	101	10	US-09-864-761-45805	Sequence 45805, A
39	38	41.3	178	9	US-10-131-241-60	Sequence 60, Appl
40	38	41.3	179	9	US-10-131-241-57	Sequence 57, Appl
41	38	41.3	180	9	US-10-131-241-47	Sequence 47, Appl
42	38	41.3	180	9	US-10-131-241-56	Sequence 56, Appl
43	38	41.3	181	9	US-10-131-241-55	Sequence 55, Appl
44	38	41.3	182	9	US-10-131-241-54	Sequence 54, Appl
45	38	41.3	182	10	US-09-998-831-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1  
US-10-016-725-16  
; Sequence 16, Application US/10016725  
; Patent No. US20020151018A1  
; GENERAL INFORMATION:  
; APPLICANT: Prouty, Stephen  
; APPLICANT: Zhang, Lin  
; APPLICANT: Stenn, Kurt  
; TITLE OF INVENTION: Stearyl-CoA Desaturase Promoter  
; FILE REFERENCE: JAJ2065  
; CURRENT APPLICATION NUMBER: US/10/016,725  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 16  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-016-725-16

Query Match 100.0%; Score 92; DB 12; Length 16;  
Best Local Similarity 100.0%; Pred. No. 6.8e-09;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPAMP 16  
| | | | | | | | | | | | | | | | | |  
Db 1 MPAHLQDDVSFPAMP 16

RESULT 2  
US-10-050-036-179  
; Sequence 179, Application US/10060036  
; Publication No. US20030073144A1  
; GENERAL INFORMATION:  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Persing, David H.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Jiang, Yuqin  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; AND DIAGNOSIS OF PANCREATIC CANCER

; FILE REFERENCE: 210121.566  
; CURRENT APPLICATION NUMBER: US/10/060.036  
; CURRENT FILING DATE: 2002-01-30  
; NUMBER OF SEQ ID NOS: 4560  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 179  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-060-036-179

Query Match 60.9%; Score 56; DB 9; Length 359;  
Best Local Similarity 90.9%; Pred. No. 0.2;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11  
||| ||||| :||  
Db 1 MPAHLQDDIS 11

RESULT 3  
US-10-029-654-12  
; Sequence 12, Application US/10029654  
; Patent No. US20020150958A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim Pharma KG  
; TITLE OF INVENTION: Methods for identifying substances for treating  
; TITLE OF INVENTION: inflammatory conditions  
; FILE REFERENCE: 1/1178  
; CURRENT APPLICATION NUMBER: US/10/029,654  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: US 60/257,878  
; PRIOR FILING DATE: 2000-12-22  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-029-654-12

Query Match 60.9%; Score 56; DB 12; Length 359;  
Best Local Similarity 90.9%; Pred. No. 0.2;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11  
||| ||||| :||  
Db 1 MPAHLQDDIS 11

RESULT 4  
US-10-001-857-179  
; Sequence 179, Application US/10001857  
; Publication No. US20020183500A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro  
; FILE REFERENCE: DEX-0273  
; CURRENT APPLICATION NUMBER: US/10/001,857  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/252,054  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 208  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 179  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-001-857-179

Query Match 47.8%; Score 44; DB 9; Length 71;  
Best Local Similarity 61.5%; Pred. No. 3;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 4 HLLQDDVSFPAPW 16  
|| | || | :||  
Db 38 HLSGDRVSTPSPW 50

RESULT 5  
US-09-971-536-42  
; Sequence 42, Application US/09971536  
; Patent No. US20020159976A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka  
; APPLICANT: Bloksberg, Leonard  
; APPLICANT: Lubbers, Mark  
; APPLICANT: Dekker, James  
; APPLICANT: Christenson, Anna  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul  
; APPLICANT: Reid, Julian  
; APPLICANT: Coolbear, Timothy  
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth  
; TITLE OF INVENTION: Using Them  
; FILE REFERENCE: 1043c2  
; CURRENT APPLICATION NUMBER: US/09/971,536  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238  
; PRIOR FILING DATE: 2000-08-08  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160  
; PRIOR FILING DATE: 2001-08-08  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 42  
; LENGTH: 632  
; TYPE: PRT  
; ORGANISM: Lactobacillus rhamnosus  
US-09-971-536-42

Query Match 47.3%; Score 43.5; DB 9; Length 632;  
Best Local Similarity 42.9%; Pred. No. 44;  
Matches 9; Conservative 4; Mismatches 1; Indels 7; Gaps 1;

QY 1 MPAHL-----LQDDVSFPA 14  
|| | | :|| :|| :||  
Db 430 MPGLVNASYDPLKNDITEPA 450

RESULT 6  
US-10-131-241-49  
; Sequence 49, Application US/10131241  
; Publication No. US20030012792A1  
; GENERAL INFORMATION:  
; APPLICANT: Holaday, John W.  
; APPLICANT: Fortier, Anne H.  
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prol  
; FILE REFERENCE: 05213-0344 43170-271565  
; CURRENT APPLICATION NUMBER: US/10/131,241  
; CURRENT FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: US 09/413,049  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: US 09/316,802  
; PRIOR FILING DATE: 1999-05-21  
; PRIOR APPLICATION NUMBER: US 60/086,586  
; PRIOR FILING DATE: 1998-05-22  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: PatentIn version 3.1

```

RESULT 8
US-10-228-931-4
; Sequence 4, Application US/10228931
; Publication No. US20030051258A1
; GENERAL INFORMATION:
; APPLICANT: Verma, Ajit K
; APPLICANT: Reddig, Peter J
; APPLICANT: Jansen, Aaron P
; TITLE OF INVENTION: Animal Model System for Squamous Cell Carcinoma
; FILE REFERENCE: 960296, 97613
; CURRENT APPLICATION NUMBER: US/10/228, 931
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: US/09/772, 647
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

```

RESULT 10
US-09-738-626-6569
; Sequence 6569, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENO, AKIHO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDE
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377481
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159161
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059

```



; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 6569  
; LENGTH: 81  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6569

Query Match 43.5%; Score 40; DB 9; Length 81;  
Best Local Similarity 35.7%; Pred. No. 16;  
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSPFA 14  
Db 7 LFNHLHRTDITPS 20

RESULT 11  
US-09-764-891-2789  
; Sequence 2789, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 2789  
; LENGTH: 128  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-891-2789

Query Match 43.5%; Score 40; DB 9; Length 128;  
Best Local Similarity 54.5%; Pred. No. 27;  
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 LODDVSPFPAW 16  
Db 12 LSPETGFPAPW 22

RESULT 12  
US-10-224-446-30  
; Sequence 30, Application US/10224446  
; Publication No. US20030024010A1  
; GENERAL INFORMATION:  
; APPLICANT: CAHOON, EDGAR B.  
; APPLICANT: CARLSON, THOMAS J.  
; APPLICANT: HITZ, WILLIAM D.  
; APPLICANT: RIPP, KEVIN G.  
; TITLE OF INVENTION: GENES FOR PLANT PATTY ACID MODIFYING ENZYMES  
; TITLE OF INVENTION: ASSOCIATED  
; TITLE OF INVENTION: WITH CONJUGATED DOUBLE BOND FORMATION  
; FILE REFERENCE: BB-1249  
; CURRENT APPLICATION NUMBER: US/10/224,446  
; CURRENT FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: US/09/375,299  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/097,186  
; PRIOR FILING DATE: August 20, 1998  
; PRIOR FILING DATE: July 8, 1999  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: MICROSOFT OFFICE 97  
; SEQ ID NO 30  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Aleurites fordii  
US-10-224-446-30

Query Match 43.5%; Score 40; DB 9; Length 255;  
Best Local Similarity 46.2%; Pred. No. 58;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 HLLQDDVSPFPAW 16  
Db 38 HLLPSPITYIAMP 50

RESULT 13  
US-09-864-761-42012  
; Sequence 42012, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 42012  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL137190.2  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7

```

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.3
; OTHER INFORMATION: SWISSPROT HIT: O84337, EVALUE 4.70e+00
; OTHER INFORMATION: EST_HUMAN HIT: R78762.1, EVALUE 2.40e+00
US-09-864-761-42012

```

```

Query Match          42.9%; Score 39.5; DB 10; Length 72;
Best Local Similarity 44.4%; Pred. No. 17;
Matches 8; Conservative 3; Mismatches 4; Indels 4; Gaps 1;

```

```

QY 2 PAHL--LQDDVSFPAPW 16
    ||| |||:| ||
Db 38 PSHLKKPLQDEIPRGWEP 55

```

```

RESULT 14
US-09-910-009A-56
; Sequence 56, Application US/09910009A
; Publication No. US20030050234A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Research Foundation
; APPLICANT: Cognetix, Inc.
; APPLICANT: Olivera, Baldomero M.
; APPLICANT: McIntosh, J. Michael
; APPLICANT: Garrett, James E.
; APPLICANT: Watkins, Maren
; APPLICANT: Cruz, Lourdes J.
; APPLICANT: Shon, Ki-Joon
; APPLICANT: Jacobsen, Richard
; APPLICANT: Jones, Robert M.
; APPLICANT: Cartier, G. Edward
; APPLICANT: Shen, Greg S.
; APPLICANT: Wagstaff, John D.
; TITLE OF INVENTION: Mu-Conopeptides
; FILE REFERENCE: 2314-242
; CURRENT APPLICATION NUMBER: US/09/910,009A
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 60/219,619
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/245,157
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/264,319
; PRIOR FILING DATE: 2001-01-29
; PRIOR APPLICATION NUMBER: US 60/277,270
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 520
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Conus characteristicus
US-09-910-009A-56

```

```

Query Match          42.4%; Score 39; DB 9; Length 74;
Best Local Similarity 46.7%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 5; Indels 5; Gaps 0;

```

```

QY 2 PAHLQDDVSFPAPW 16
    || :|||:| :|
Db 32 PADRMQDDISSEQYP 46

```

```

RESULT 15
US-09-910-009A-74
; Sequence 74, Application US/09910009A
; Publication No. US20030050234A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Research Foundation
; APPLICANT: Cognetix, Inc.
; APPLICANT: Olivera, Baldomero M.
; APPLICANT: McIntosh, J. Michael
; APPLICANT: Garrett, James E.
; APPLICANT: Watkins, Maren

```

```

; APPLICANT: Cruz, Lourdes J.
; APPLICANT: Shon, Ki-Joon
; APPLICANT: Jacobsen, Richard
; APPLICANT: Jones, Robert M.
; APPLICANT: Cartier, G. Edward
; APPLICANT: Shen, Greg S.
; APPLICANT: Wagstaff, John D.
; TITLE OF INVENTION: Mu-Conopeptides
; FILE REFERENCE: 2314-242
; CURRENT APPLICATION NUMBER: US/09/910,009A
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 60/219,619
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/245,157
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/264,319
; PRIOR FILING DATE: 2001-01-29
; PRIOR APPLICATION NUMBER: US 60/277,270
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 520
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 74
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Conus circumciscus
US-09-910-009A-74

```

```

Query Match          42.4%; Score 39; DB 9; Length 74;
Best Local Similarity 46.7%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 5; Indels 5; Gaps 0;

```

```

QY 2 PAHLQDDVSFPAPW 16
    || :|||:| :|
Db 32 PADRMQDDISSEQYP 46

```

```

Search completed: May 14, 2003, 15:30:19
Job time : 26 secs

```

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:30:26 ; Search time 44 seconds  
(without alignments)  
34.958 Million cell updates/sec

Title: US-10-016-725-16

Perfect score: 92  
Sequence: 1 MPAHLQDDVSFPWP 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	48.9	1675	1 LRTH	clathrin heavy cha
2	44	47.8	1687	2 T43144	vitellogenin II pr
3	43.5	47.3	518	2 T01318	pectinesterase hom
4	43	46.7	136	2 S72870	hypothetical prote
5	43	46.7	414	2 E84653	probable poly(A) p
6	42	45.7	663	1 A35087	arachidonate 12-li
7	42	45.7	1280	2 T51500	hypothetical prote
8	41.5	45.1	245	2 S39481	glutamate-ammonia
9	41.5	45.1	357	2 S39477	glutamate-ammonia
10	41.5	45.1	627	2 H82573	exonuclease ABC s
11	41	44.6	160	2 A84104	hypothetical prote
12	41	44.6	227	2 A70885	hypothetical prote
13	41	44.6	338	2 T21161	hypothetical prote
14	41	44.6	387	2 F82746	DNA processing cha
15	41	44.6	547	2 T22856	hypothetical prote
16	41	44.6	642	2 A82715	hypothetical prote
17	41	44.6	660	2 A00661	exonuclease ABC c
18	41	44.6	681	2 C97497	invasin-like prote
19	41	44.6	685	2 A63408	exonuclease ABC c
20	41	44.6	707	1 A53530	exonuclease ABC c
21	41	44.6	736	1 KIRBCE	protein kinase C (
22	41	44.6	737	1 S28942	protein kinase C (
23	41	44.6	737	1 KIRTCE	protein kinase C (
24	41	44.6	737	1 KIMSCE	protein kinase C (
25	41	44.6	847	2 G85666	hypothetical prote
26	41	44.6	847	2 A80636	periplasmic glucan
27	41	44.6	847	2 F84847	glucan biosynthesi
28	41	44.6	857	2 C90807	membrane glycosyl
29	41	44.6	1146	2 F84487	probable ABC trans

30	41	44.6	2535	2 T02646	hypothetical prote
31	40.5	44.0	356	2 T03253	glutamate-ammonia
32	40.5	44.0	844	2 S77383	hypothetical prote
33	40	43.5	175	1 G69518	isochorismatase (e
34	40	43.5	243	1 JJAG32	trans-zeatin secre
35	40	43.5	243	2 S03122	trans-zeatin secre
36	40	43.5	243	2 AB3248	trans-zeatin secre
37	40	43.5	320	2 E82520	hypothetical prote
38	40	43.5	377	2 T04585	hypothetical prote
39	40	43.5	522	2 G96526	hypothetical prote
40	40	43.5	1042	2 S76045	hypothetical prote
41	40	43.5	1089	2 S53978	p8E1 protein - yea
42	40	43.5	1102	2 H84545	probable ubiquitin
43	39.5	42.9	374	2 F87596	hypothetical prote
44	39	42.4	108	2 B84229	hypothetical prote
45	39	42.4	130	2 T08329	hypothetical prote

## ALIGNMENTS

RESULT 1  
LRTH  
clathrin heavy chain - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 22-Jun-1999  
C:Accession: A39941  
R:Kirchhausen, T.; Harrison, S.C.; Ping Chow, E.; Mattaliano, R.J.; Ramachandran, K.; Proc. Natl. Acad. Sci. U.S.A. 84, 8805-8809, 1987  
A:Title: Clathrin heavy chain: molecular cloning and complete primary structure.  
A:Reference number: A39941; MUID:88057376; PMID:3480512  
A:Accession: A39941  
A:Molecule type: mRNA  
A:Residues: 1-1675 <KIR>  
A:Cross-references: GB:J03583; NID:g203301; PIDN:AAA40874.1; PID:g203302  
C:Comment: Clathrin, the major protein component of coated pits and vesicles, is a t r its carboxyl end. The heavy chains are also held together by noncovalent interacti  
C:Comment: The amino end of the mature protein is blocked.  
C:Superfamily: clathrin heavy chain  
C:Keywords: coated pits  
F:1-479/Domain: amino-terminal <TER>  
F:480-523/Region: link  
F:524-634/Domain: distal <DIS>  
F:635-638/Region: joint #status predicted  
F:639-1675/Domain: proximal <PRX>

Query Match 48.9%; Score 45; DB 1; Length 1675;  
Best Local Similarity 53.3%; Pred. No. 43;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPWP 15  
| | | | | | | | | |  
Db 99 MKAHMTDDVTFWKW 113

## RESULT 2

T43144  
vitellogenin II precursor - mummichog  
C:Species: Fundulus heteroclitus (mummichog)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
C:Accession: T43144  
R:Laflaur, G.J.; Hoch, K.L.; Denslow, N.; Byrne, B.M.; Wallace, R.A.  
submitted to the EMBL Data Library, September 1996  
A:Description: Derivation of oocyte and egg proteins from parental vitellogenins in l  
n.  
A:Reference number: Z22316  
A:Accession: T43144  
A>Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-1687 <LAF>  
A:Cross-references: EMBL:U70826; NID:g1621358; PIDN:AAB17152.1  
A:Experimental source: estrogen-induced liver  
C:Superfamily: vitellogenin

C:Keywords: egg yolk; phosphoprotein  
 F:1-15/Domain: signal sequence status predicted <Sig>  
 F:16-1687/Product: vitellogenin II #status predicted <MAT>

Query Match 47.8%; Score 44; DB 2; Length 1687;  
 Best Local Similarity 61.5%; Pred. No. 64;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFPAW 14  
 |||||: 1:1 |  
 DB 831 PAHLKSDISMKA 843

## RESULT 3

T01318 pectinesterase homolog T14P8.1 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 24-Sep-1999

C:Accession: T01318

R:Kalicki, J.; Elliott, G.; Cloud, J.

submitted to the EMBL Data Library, May 1998

A:Description: The sequence of A. thaliana T14P8.

A:Reference number: Z14290

A:Accession: T01318

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-518 <KAL>

A:Cross-references: EMBL:AF069298; NID:g3193282; PIDN:AAC19272.1; PID:g3193288

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 235/2; 289/1

A:Note: T14P8.1

C:Superfamily: pectinesterase

Query Match 47.3%; Score 43.5; DB 2; Length 518;  
 Best Local Similarity 38.9%; Pred. No. 20;  
 Matches 7; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 MPAHL---LQDDVSFPAW 15  
 :| | :||| |  
 DB 171 IPGHPKGVKEDVGPPMW 188

## RESULT 4

S72870

hypothetical protein B2126\_F1\_18 - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 23-Mar-2001

C:Accession: S72870

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid B2126.

A:Reference number: S72585

A:Accession: S72870

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-136 <SMI>

A:Cross-references: EMBL:U00017; NID:g466994; PIDN:AAA17210.1; PID:g467025

C:Superfamily: Mycobacterium leprae hypothetical protein B2126\_F1\_18

Query Match 46.7%; Score 43; DB 2; Length 136;  
 Best Local Similarity 70.0%; Pred. No. 5.1;  
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAHLQDDVS 11  
 |||||: 1:1 |  
 DB 36 PAHVLQDDLT 45

## RESULT 5

E84653

probable poly(A) polymerase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: E84653

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
 Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: E84653

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-414 <STO>

A:Cross-references: GB:AE002093; NID:g3643598; PIDN:AAC42245.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g25850

A:Map position: 2

Query Match 46.7%; Score 43; DB 2; Length 414;  
 Best Local Similarity 66.7%; Pred. No. 19;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 7 QDDVSFPAW 15  
 :|||: ||| |  
 DB 295 EDDLSPVW 303

## RESULT 6

A35087

arachidonate 12-lipoxygenase (EC 1.13.11.31) - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: A35087

R:Yoshimoto, T.; Suzuki, H.; Yamamoto, S.; Takai, T.; Yokoyama, C.; Tanabe, T.

Proc. Natl. Acad. Sci. U.S.A. 87, 2142-2146, 1990

A:Title: Cloning and sequence analysis of the cDNA for arachidonate 12-lipoxygenase

A:Reference number: A35087; MUID:90192763; PMID:2315307

A:Accession: A35087

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-663 <YOS>

A:Cross-references: GB:M31417

C:Superfamily: arachidonate 5-lipoxygenase

C:Keywords: oxidoreductase

Query Match 45.7%; Score 42; DB 1; Length 663;  
 Best Local Similarity 66.7%; Pred. No. 48;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 HLLQDDVSFPAW 15  
 |||||: |  
 DB 59 HLLQDDAWFCNW 80

## RESULT 7

T51500

hypothetical protein F5E19\_20 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000

C:Accession: T51500

R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.;

submitted to the Protein Sequence Database, August 2000

A:Reference number: Z25394

A:Accession: T51500

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1280 <SAT>

A:Cross-references: EMBL:AL391147

A:Experimental source: cultivar Columbia; BAC clone F5E19

C:Genetics:

A:Map position: 5

A:Introns: 272/3; 301/2; 331/1; 425/1; 557/1; 808/2; 885/2; 927/1; 1019/3

A:Note: F5E19\_20

Query Match 45.7%; Score 42; DB 2; Length 1280;  
 Best Local Similarity 50.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSPPA 14  
 DB 765 IFSRLVEPVSPPA 778  
 ||:|:::|||||

RESULT 8  
 S39481  
 glutamate-ammonia ligase (EC 6.3.1.2) 1-5, cytosolic - maize (fragment)  
 N:Alternate names: glutamine synthetase  
 C:Species: Zea mays (maize)  
 C:Date: 07-Apr-1994 #sequence\_revision 08-Nov-1996 #text\_change 03-Jun-2002  
 C:Accession: S39481  
 R:Li, M.G.; Villemur, R.; Hussey, P.J.; Silflow, C.D.; Gantt, J.S.; Snustad, D.P.  
 Plant Mol. Biol. 23, 401-407, 1993  
 A:Title: Differential expression of six glutamine synthetase genes in Zea mays.  
 A:Reference number: S39477; MUID:94033318; PMID:8106013  
 A:Accession: S39481  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-245 <LIM>  
 A:Cross-references: EMBL:X65930; NID:9434331; PIDN:CAA46723.1; PID:gl360700  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992  
 C:Superfamily: glutamate-ammonia ligase  
 C:Keywords: cytosol; ligase

Query Match 45.1%; Score 41.5; DB 2; Length 245;  
 Best Local Similarity 69.2%; Pred. No. 18;  
 Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 5 LLQDDVSFP-AMP 16  
 DB 22 LLQKDVSWPLGWP 34  
 |||||:|:|

RESULT 9  
 S39477  
 glutamate-ammonia ligase (EC 6.3.1.2) 1-1, cytosolic - maize  
 N:Alternate names: glutamine synthetase  
 C:Species: Zea mays (maize)  
 C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 03-Jun-2002  
 C:Accession: S39477  
 R:Li, M.G.; Villemur, R.; Hussey, P.J.; Silflow, C.D.; Gantt, J.S.; Snustad, D.P.  
 Plant Mol. Biol. 23, 401-407, 1993  
 A:Title: Differential expression of six glutamine synthetase genes in Zea mays.  
 A:Reference number: S39477; MUID:94033318; PMID:8106013  
 A:Accession: S39477  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-357 <LIM>  
 A:Cross-references: EMBL:X65926; NID:9434323; PIDN:CAA46719.1; PID:g434324  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992  
 C:Superfamily: glutamate-ammonia ligase  
 C:Keywords: cytosol; ligase

Query Match 45.1%; Score 41.5; DB 2; Length 357;  
 Best Local Similarity 69.2%; Pred. No. 28;  
 Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 5 LLQDDVSFP-AMP 16  
 DB 134 LLQKDVSWPLGWP 146  
 |||||:|:|

RESULT 10  
 H82573  
 excinuclease ABC subunit C XF2311 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: H82573  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: H82573  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-627 <SIM>  
 A:Cross-references: GB:AE004042; GB:AE003849; NID:g9107469; PIDN:AAF85110.1; GSPDB:GI  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrei  
 as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; F  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; L  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martini  
 A:Authors: Martins, E.M.F.; Matsukuma, A.V.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
 A:Reference number: A59328  
 A:Contents: annotation  
 A:Genetics:  
 C:Gene: XF2311  
 C:Superfamily: excinuclease ABC chain C

Query Match 45.1%; Score 41.5; DB 2; Length 627;  
 Best Local Similarity 47.4%; Pred. No. 54;  
 Matches 9; Conservative 2; Mismatches 1; Indels 7; Gaps 1;

QY 5 LLQDDVSFP-----ANP 16  
 DB 105 LLRDDKSPYVLLTREANP 123  
 |||||:|:|

RESULT 11  
 A84104  
 hypothetical protein BH3633 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: A84104  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; F  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans  
 A:Reference number: A83650; MUID:20512582; PMID:11058132  
 A:Accession: A84104  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-160 <STO>  
 A:Cross-references: GB:AP001519; GB:BA000004; NID:gl0176109; PIDN:BAB07352.1; GSPDB:C  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH3633

Query Match 44.6%; Score 41; DB 2; Length 160;  
 Best Local Similarity 77.8%; Pred. No. 14;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 HLLQDDVSF 12  
 DB 50 HLQDDLSF 58  
 ||:|:|:|

RESULT 12  
 A70885  
 hypothetical protein Rv2794c - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: A70885

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230  
 A: Accession: T21161  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-227 <COL>  
 A: Cross-references: GB:AL008967; GB:AL123456; NID: g3261491; PIDN: CAAL5589.1; PID: g262431  
 A: Experimental source: strain H37RV  
 C: Genetics:  
 A: Gene: Rv2794c

Query Match 44.6%; Score 41; DB 2; Length 227;  
 Best Local Similarity 46.7%; Pred. No. 20;  
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFPWP 16  
 |||::|::|::|  
 Db 70 PAPILKGDKEPCWP 84

## RESULT 13

T21161

hypothetical protein F20E11.10 - Caenorhabditis elegans

C: Species: Caenorhabditis elegans

C: Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C: Accession: T21161

R: Ainscough, R.

submitted to the EMBL Data Library, November 1996

A: Reference number: F21161

A: Accession: T21161

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: DNA

A: Residues: 1-338 &lt;WIL&gt;

A: Cross-references: EMBL: Z81508; PIDN: CAB04147.1; GSPDB: GN00023; CESP: F20E11.10

A: Experimental source: clone F20E11

C: Genetics:

A: Gene: CESP: F20E11.10

A: Map position: 5

A: Introns: 110/1; 173/3

C: Superfamily: Caenorhabditis hypothetical protein C49G7.2

Query Match 44.6%; Score 41; DB 2; Length 338;  
 Best Local Similarity 50.0%; Pred. No. 32;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFP 13  
 |::|::|::|  
 Db 186 PVYIADDTVP 197

## RESULT 14

F82746

DNA processing chain A XF0924 [imported] - Xylella fastidiosa (strain 9a5c)

C: Species: Xylella fastidiosa

C: Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C: Accession: F82746

R: anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A: Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A: Reference number: A82515; MUID: 20365717; PMID: 10510347

A: Note: for a complete list of authors see reference number A59328 below

A: Accession: F82746

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-387 &lt;SIM&gt;

A: Cross-references: GB: AE003931; GB: AE003849; NID: g9105834; PIDN: AAF83734.1; GSPDB: GN001

A: Experimental source: strain 9a5c

R: Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A: Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima J.P.; Krieger, J.E.; Kuramae, E.B.; La  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.D.; Marques, M.V.; Martins  
 A: Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
 M.; Tsubako, M.H.; Vallada, H.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
 A: Reference number: A59328  
 A: Contents: annotation  
 C: Genetics:  
 A: Gene: XF0924

Query Match 44.6%; Score 41; DB 2; Length 387;  
 Best Local Similarity 42.9%; Pred. No. 38;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSFPWP 16  
 |||::|::|::|  
 Db 307 AHTLRERLDAPSWP 320

## RESULT 15

T22856

hypothetical protein F57F5.5 - Caenorhabditis elegans

C: Species: Caenorhabditis elegans

C: Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Jun-2002

C: Accession: T22856

R: Harris, B.

submitted to the EMBL Data Library, July 1996

A: Reference number: Z19627

A: Accession: T22856

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: DNA

A: Residues: 1-547 &lt;WIL&gt;

A: Cross-references: EMBL: Z75953; PIDN: CAB00101.1; GSPDB: GN00023; CESP: F57F5.5

A: Experimental source: clone F57F5

C: Genetics:

A: Gene: CESP: F57F5.5

A: Map position: 5

A: Introns: 43/1; 129/2; 166/3; 187/1; 289/3; 341/2; 450/3; 499/3

F: 11-60/Domain: protein kinase C zinc-binding repeat homology &lt;KZN1&gt;

F: 89-139/Domain: protein kinase C zinc-binding repeat homology &lt;KZN2&gt;

Query Match 44.6%; Score 41; DB 2; Length 547;  
 Best Local Similarity 54.5%; Pred. No. 57;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLQDDVSFPWP 15  
 :|||::|::|  
 Db 429 ILNDDVLYPVW 439

Search completed: May 14, 2003, 15:32:58

Job time : 46 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:24:22 ; Search time 11 Seconds  
(without alignments)  
60.329 Million cell updates/sec

Title: US-10-016-725-16

Perfect score: 92

Sequence: 1 MPAHLQDVSPAMP 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	60.9	359	1 ACOD_HUMAN	O00767 homo sapien
2	48	52.2	359	1 ACOD_SHEEP	O62849 ovis aries
3	46	50.0	359	1 ACOD_BOVIN	Q9tt94 bos taurus
4	45	48.9	1675	1 CLH1_HUMAN	Q00610 homo sapien
5	45	48.9	1675	1 CLH1_BOVIN	P49951 bos taurus
6	45	48.9	1675	1 CLH1_RAT	P11442 rattus norv
7	44	47.8	1687	1 VIT2_FUNHE	Q98893 fundulus he
8	42	45.7	662	1 LOX2_PIG	P16469 sus scrofa
9	41.5	45.1	357	1 GLN1_MAIZE	P38559 zea mays (m
10	41	44.6	707	1 KPCE2_CABEL	P34885 caenorhabdi
11	41	44.6	736	1 KPCE_RABIT	P10830 oryctolagus
12	41	44.6	737	1 KPCE_HUMAN	Q02156 homo sapien
13	41	44.6	737	1 KPCE_MOUSE	P16034 mus musculu
14	41	44.6	737	1 KPCE_RAT	P09216 rattus norv
15	41	44.6	743	1 KPCE2_APLICA	Q16975 aplysia cal
16	41	44.6	847	1 MDOH_ECOLI	P33137 escherichia
17	40	43.5	243	1 IPT2_AGR75	P58758 agrobacteri
18	40	43.5	243	1 IPT2_AGR77	P06524 agrobacteri
19	40	43.5	246	1 MPCL_CYNDA	O04701 cynodon dac
20	40	43.5	495	1 SR52_ARATH	P49966 arabidopsis
21	40	43.5	495	1 SR52_ARATH	P49967 arabidopsis
22	40	43.5	583	1 HEMO_BRARE	Q9yht4 brachydanio
23	40	43.5	1089	1 IMB3_YEAST	P32337 saccharomyc
24	39.5	42.9	389	1 5H1B_CAVPO	O08892 cavia porce
25	39	42.4	308	1 YPX1_BLVJ	P03412 bovine leuk
26	39	42.4	354	1 ACOD_MESAU	Q64420 mesocricetu
27	39	42.4	355	1 NDFL_MESAU	Q60430 mesocricetu
28	39	42.4	356	1 NDFL_HUMAN	Q13562 homo sapien
29	39	42.4	357	1 NDFL_MOUSE	Q60867 mus musculu
30	39	42.4	357	1 NDFL_RAT	Q64289 rattus norv
31	39	42.4	412	1 DOK2_MOUSE	O70469 mus musculu
32	39	42.4	495	1 HYDA_PSEPU	Q59699 pseudomonas
33	39	42.4	498	1 SYNC_DEIRA	Q9rtt6 deinococcus

34 39 42.4 634 1 KPC3\_DROME  
35 39 42.4 958 1 YGX7\_YEAST  
36 39 42.4 1032 1 YGFK\_ECO57  
37 39 42.4 1032 1 YGFK\_ECOLI  
38 39 42.4 1293 1 ENTE\_ECO57  
39 39 42.4 1293 1 ENTE\_ECOLI  
40 39 42.4 1678 1 CLH1\_DROME  
41 39 42.4 3432 1 POLG\_JAEV1  
42 39 42.4 3432 1 POLG\_JAEV5  
43 39 42.4 3432 1 POLG\_JAEVJ  
44 39 42.4 5217 1 HTSL\_COCCA  
45 38.5 41.8 356 1 GLN2\_PNAVU

#### ALIGNMENTS

RESULT 1  
ACOD\_HUMAN STANDARD; PRT; 359 AA.  
AC O00767; Q16150; Q9Y695; Q9BS07; Q96KF6;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Acyl-CoA desaturase (EC 1.14.99.5) (Stearoyl-CoA desaturase) (Fatty  
DE acid desaturase) (Delta(9)-desaturase).  
GN SCD.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Al-Jeryan L., McCord A., Pierotti A.R., Craft J.A.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver, Brain, and Skin;  
RX MEDLINE=99247918; PubMed=10229681;  
RA Zhang L., Ge L., Parimoo S., Stenn K., Prouty S.M.;  
RT "Human stearoyl-CoA desaturase: alternative transcripts generated from  
RL a single gene by usage of tandem polyadenylation sites.";  
RN Biochem. J. 340:255-264(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Hoshino T., Ohtsu K.;  
RT "Cloning, sequencing and expression of human stearoyl-CoA  
RL desaturase.";  
RN Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 3-239 FROM N.A.  
RC TISSUE=Adipose tissue;  
RX MEDLINE=94222609; PubMed=7909540;  
RA Li J., Ding S.-F., Habib N.A., Ferner B.F., Wood C.B., Gilmour R.S.;  
RT "Partial characterization of a cDNA for human stearoyl-CoA desaturase  
RL and changes in its mRNA expression in some normal and malignant  
RN tissues.";  
RL Int. J. Cancer 57:348-352(1994).  
RN [6]  
RP SEQUENCE OF 1-9 FROM N.A.  
RA Zhang L., Ge L., Tran T., Stenn K., Prouty S.M.;  
RT "Isolation and characterization of the human stearoyl-CoA desaturase  
RL gene promoter: requirement of a conserved CCAAT cis-element.";  
RN Biochem. J. 0:0:0(2001).  
CC -!- FUNCTION: TERMINAL COMPONENT OF THE LIVER MICROSOMAL STEARYL-COA  
CC DESATURASE SYSTEM, THAT UTILIZES O(2) AND ELECTRONS FROM REDUCED  
CC CYTOCHROME B(5) TO CATALYZE THE INSERTION OF A DOUBLE BOND INTO A

CC SPECTRUM OF FATTY ACYL-CoA SUBSTRATES AMONG WHICH PALMITOYL-CoA  
 CC AND STEAROYL-CoA.  
 CC -!- CATALYTIC ACTIVITY: Stearoyl-CoA + AH(2) + O(2) = oleoyl-CoA + A +  
 CC 2 H(2)O.  
 CC -!- COFACTOR: IRON.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum (Probable).  
 CC -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE  
 CC AND/OR BE INVOLVED IN METAL ION BINDING.  
 CC -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC  
 CC EMBL; Y13647; CAA73998.1; -  
 CC EMBL; AF097514; AAD29870.1; -  
 CC EMBL; AB032261; BAA93510.1; -  
 CC EMBL; BC005807; AAB05807.1; -  
 CC EMBL; S70284; AAB30631.1; -  
 CC EMBL; AF320307; AAK54510.1; -  
 CC Genew; HGNC:10571; SCD.  
 CC MIM; 604031; -  
 CC InterPro; IPR001522; Desaturase.  
 CC InterPro; IPR001225; FA\_desaturase.  
 CC Pfam; PF00487; FA\_desaturase; 1.  
 CC PRINTS; PR00075; FACDDSATRASE.  
 CC ProDom; PD002221; Desaturase; 1.  
 CC PROSITE; PS00476; FATTY\_ACID\_DESATUR\_1; 1.  
 CC Oxidoreductase; Fatty acid biosynthesis; Transmembrane;  
 CC Endoplasmic reticulum; Iron.  
 CC TRANSMEM 76 96 POTENTIAL.  
 CC TRANSMEM 98 118 POTENTIAL.  
 CC TRANSMEM 223 243 POTENTIAL.  
 CC TRANSMEM 315 335 POTENTIAL.  
 CC DOMAIN 120 125 HISTIDINE BOX-1.  
 CC DOMAIN 157 161 HISTIDINE BOX-2.  
 CC DOMAIN 298 302 HISTIDINE BOX-3.  
 CC CONFLICT 5 5 L -> M (IN REF. 5).  
 CC CONFLICT 8 8 D -> E (IN REF. 5).  
 CC CONFLICT 25 26 SR -> PG (IN REF. 1 AND 5).  
 CC CONFLICT 224 224 M -> L (IN REF. 2 AND 3).  
 CC CONFLICT 237 237 F -> C (IN REF. 5).  
 CC CONFLICT 320 320 T -> N (IN REF. 1).  
 CC CONFLICT 326 326 C -> W (IN REF. 1).  
 CC CONFLICT 333 333 A -> T (IN REF. 1).  
 CC CONFLICT 356 359 MISSING (IN REF. 4).  
 CC SEQUENCE 359 AA; 41522 MW; ED56A63BDB850F05 CRC64;  
 CC  
 CC Query Match 60.9%; Score 56; DB 1; Length 359;  
 CC Best Local Similarity 90.9%; Pred. No. 0.041;  
 CC Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 MPAHLQDDVS 11  
 CC |  
 CC Db 1 MPAHLQDDIS 11  
 CC |  
 CC RESULT 2  
 CC ACOD\_SHEEP STANDARD; PRT; 359 AA.  
 CC AC 062849;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Acyl-CoA desaturase (EC 1.14.99.5) (Stearoyl-CoA desaturase) (Fatty  
 CC acid desaturase) (delta(9)-desaturase).  
 CC GN SCD.  
 CC OS Ovis aries (Sheep).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Caprinae; Ovis.  
 CC NCBI\_TaxID=9940;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=Y1089; TISSUE=Adipose tissue;  
 CC RX MEDLINE=98223428; PubMed=9554990;  
 CC RA Ward R.J., Travers M.T., Richards S.E., Vernon R.G., Salter A.M.,  
 CC Buttery P.J., Barber M.C.;  
 CC "Stearoyl-CoA desaturase mRNA is transcribed from a single gene in the  
 CC ovine genome.";  
 CC Biochim. Biophys. Acta 1391:145-156(1998).  
 CC -!- FUNCTION: TERMINAL COMPONENT OF THE LIVER MICROSOMAL STEAROYL-CoA  
 CC DESATURASE SYSTEM, THAT UTILIZES O(2) AND ELECTRONS FROM REDUCED  
 CC CYTOCHROME B(5) TO CATALYZE THE INSERTION OF A DOUBLE BOND INTO A  
 CC SPECTRUM OF FATTY ACYL-CoA SUBSTRATES AMONG WHICH PALMITOYL-CoA  
 CC AND STEAROYL-CoA (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Stearoyl-CoA + AH(2) + O(2) = oleoyl-CoA + A +  
 CC 2 H(2)O.  
 CC -!- COFACTOR: IRON.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum (Probable).  
 CC -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE  
 CC AND/OR BE INVOLVED IN METAL ION BINDING.  
 CC -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC  
 CC EMBL; AJ001048; CAA04502.1; -  
 CC InterPro; IPR001522; Desaturase.  
 CC InterPro; IPR001225; FA\_desaturase.  
 CC Pfam; PF00487; FA\_desaturase; 1.  
 CC PRINTS; PR00075; FACDDSATRASE.  
 CC ProDom; PD002221; Desaturase; 1.  
 CC PROSITE; PS00476; FATTY\_ACID\_DESATUR\_1; 1.  
 CC Oxidoreductase; Fatty acid biosynthesis; Transmembrane;  
 CC Endoplasmic reticulum; Iron.  
 CC TRANSMEM 76 96 POTENTIAL.  
 CC TRANSMEM 98 118 POTENTIAL.  
 CC TRANSMEM 223 243 POTENTIAL.  
 CC TRANSMEM 315 335 POTENTIAL.  
 CC DOMAIN 120 125 HISTIDINE BOX-1.  
 CC DOMAIN 157 161 HISTIDINE BOX-2.  
 CC DOMAIN 298 302 HISTIDINE BOX-3.  
 CC SEQUENCE 359 AA; 41671 MW; 219CFEBB1E353418 CRC64;  
 CC  
 CC Query Match 52.2%; Score 48; DB 1; Length 359;  
 CC Best Local Similarity 72.7%; Pred. No. 0.95;  
 CC Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 MPAHLQDDVS 11  
 CC |  
 CC Db 1 MPAHLQDDIS 11  
 CC |  
 CC RESULT 3  
 CC ACOD\_BOVIN STANDARD; PRT; 359 AA.  
 CC ID ACOD\_BOVIN  
 CC AC Q9TT94;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Acyl-CoA desaturase (EC 1.14.99.5) (Stearoyl-CoA desaturase) (Fatty  
 CC acid desaturase) (delta(9)-desaturase).  
 CC GN SCD.  
 CC OS Bos taurus (Bovine).



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Adipose tissue;  
 RA Chung M.I., Ha S.H., Baik M.G., Choi Y.J.;  
 RT "Cloning and characterization of full-coding cDNA of bovine stearoyl  
 CoA desaturase from adipose tissue.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: TERMINAL COMPONENT OF THE LIVER MICROSOMAL STEARYL-COA  
 DESATURASE SYSTEM. THAT UTILIZES O(2) AND ELECTRONS FROM REDUCED  
 CYTOCHROME B(5) TO CATALYZE THE INSERTION OF A DOUBLE BOND INTO A  
 SPECTRUM OF FATTY ACYL-COA SUBSTRATES AMONG WHICH PALMITOYL-COA  
 AND STEAROYL-COA (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Stearoyl-CoA + AH(2) + O(2) = oleoyl-CoA + A +  
 2 H(2)O.  
 CC -!- COFACTOR: IRON.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.. Endoplasmic  
 reticulum (Probable).  
 CC -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE  
 AND/OR BE INVOLVED IN METAL ION BINDING.  
 CC -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; AF188710; RA222305.1; -;  
 DR InterPro; IPR001522; Desaturase.  
 DR Pfam; PF00487; FA\_desaturase; 1.  
 DR PRINTS; PR00075; FACDSDSTRASE.  
 DR ProDom; PD002221; Desaturase; 1.  
 DR PROSITE; PS00476; FATTY\_ACID\_DESATUR\_1; 1.  
 DR Oxidoreductase; Fatty acid biosynthesis; Transmembrane;  
 KW Endoplasmic reticulum; Iron.  
 FT TRANSMEM 76 96 POTENTIAL.  
 FT TRANSMEM 98 118 POTENTIAL.  
 FT TRANSMEM 223 243 POTENTIAL.  
 FT TRANSMEM 315 335 POTENTIAL.  
 FT DOMAIN 120 125 HISTIDINE BOX-1.  
 FT DOMAIN 157 161 HISTIDINE BOX-2.  
 FT DOMAIN 298 302 HISTIDINE BOX-3.  
 SQ SEQUENCE 359 AA; 41755 MW; 22FB69FBD1846C33 CRC64;  
 Query Match 50.0%; Score 46; DB 1; Length 359;  
 Best Local Similarity 63.6%; Pred. No. 2.1;  
 Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MPAHLQDDVS 11  
 DB 1 MPAHLQDEIS 11  
 RESULT 4  
 ID CLHL\_HUMAN STANDARD; PRT; 1675 AA.  
 AC Q00610;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Clathrin heavy chain 1 (CLH-17).  
 GN CLH1 OR CLH17 OR KIAA0034.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;

RN SEQUENCE FROM N.A.  
 RP TISSUE=Bone marrow;  
 RX MEDLINE=96051387; PubMed=7584026;  
 RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,  
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;  
 RT "Prediction of the coding sequences of unidentified human genes. I.  
 RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by  
 RT analysis of randomly sampled cDNA clones from human immature myeloid  
 RT cell line KG-1.";  
 RL DNA Res. 1:27-35(1994).  
 RN [2]  
 RP SEQUENCE OF 560-864 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=92112210; PubMed=1765375;  
 RA Dodge G.R., Kowalsky I., McBride O.W., Yi H.F., Chu M.L., Saitta B.,  
 RA Stokes D.G., Iozzo R.V.;  
 RT "Human clathrin heavy chain (CLTC): partial molecular cloning,  
 RT expression, and mapping of the gene to human chromosome 17q11-qter.";  
 RL Genomics 11:174-178(1991).  
 CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
 COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES  
 LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE  
 TRANS GOLGI NETWORK.  
 CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3  
 LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE  
 PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE  
 PH AND THE CONCENTRATION OF CALCIUM.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
 VESICLES.  
 CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.  
 CC -!- DATABASE: NAMP-Atlas Genet. Cytogenet. Oncol. Haematol.;  
 WWW="http://www.infobiogen.fr/services/chronocancer/Genes/CLTCD360.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; D21260; BAA04801.1; -;  
 DR EMBL; X55878; CAA39363.1; -;  
 DR PIR; A40573; A40573.  
 DR HSP; P11442; LBPO.  
 DR Genew; HGNC:2092; CLTC.  
 DR MIM; 118995; -;  
 DR InterPro; IPR001473; Clathrin\_propel.  
 DR InterPro; IPR000547; Clathrin\_repeat.  
 DR Pfam; PF00637; Clathrin; 7.  
 DR Pfam; PF01394; Clathrin\_propel; 7.  
 DR SMART; SM00299; CLH; 7.  
 KW Coated pits.  
 FT DOMAIN 1 479 GLOBULAR TERMINAL DOMAIN.  
 FT DOMAIN 480 523 FLEXIBLE LINKER.  
 FT DOMAIN 524 1675 HEAVY CHAIN ARM.  
 FT DOMAIN 524 634 DISTAL SEGMENT.  
 FT DOMAIN 639 1675 PROXIMAL SEGMENT.  
 FT DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,  
 INVOLVED IN LATTICE DISASSEMBLY  
 (POTENTIAL).  
 FT BINDING 1213 1522 LIGHT CHAIN (BY SIMILARITY).  
 FT DOMAIN 1550 1675 TRIMERIZATION (BY SIMILARITY).  
 FT CONFLICT 560 560 Q -> R (IN REF. 2).  
 FT CONFLICT 817 817 G -> V (IN REF. 2).  
 SQ SEQUENCE 1675 AA; 191614 MW; 6C4F2D54950079E2 CRC64;  
 Query Match 48.9%; Score 45; DB 1; Length 1675;  
 Best Local Similarity 53.3%; Pred. No. 17;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 MPAHLQDDVSFPAP 15

Db 99 MKAHTMTDDVTFWKW 113

## RESULT 5

CLH\_BOVIN STANDARD; PRT; 1675 AA.  
AC P49951;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Clathrin heavy chain.  
GN CLTC.

OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;  
RX MEDLINE=96028100; PubMed=7585943;  
RA Liu S.-H., Wong M.L., Craik C.S., Brodsky F.M.;  
RT "Regulation of clathrin assembly and trimerization defined using  
RT recombinant triskelion hubs.";  
RL Cell 83:257-267(1995).

CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES  
CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE  
CC TRANS GOLGI NETWORK.  
CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3  
CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE  
CC PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE  
CC PH AND THE CONCENTRATION OF CALCIUM.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
CC VESICLES.

CC -!- DOMAIN: THE C-TERMINAL THIRD OF THE HEAVY CHAINS FORMS THE HUB OF  
CC THE TRISKELION. THIS REGION CONTAINS THE TRIMERIZATION DOMAIN AND  
CC THE LIGHT-CHAIN BINDING DOMAIN INVOLVED IN THE ASSEMBLY OF THE  
CC CLATHRIN LATTICE.  
CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U31757; AAC48524.1; -  
DR HSSP; P11442; LBPO.  
DR InterPro; IPR001473; Clathrin\_propel.  
DR InterPro; IPR000547; Clathrin\_repeat.  
DR Pfam; PF00637; Clathrin; 7.  
DR Pfam; PF01394; Clathrin\_propel; 7.  
DR SMART; SM00299; CLH; 7.  
KW Coated pits.

FT DOMAIN 1 479 GLOBULAR TERMINAL DOMAIN.  
FT DOMAIN 480 523 FLEXIBLE LINKER.  
FT DOMAIN 524 1675 HEAVY CHAIN ARM.  
FT DOMAIN 524 634 DISTAL SEGMENT.  
FT DOMAIN 639 1675 PROXIMAL SEGMENT.  
FT DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,  
FT INVOLVED IN LATTICE DISASSEMBLY  
FT (POTENTIAL).  
FT BINDING 1213 1522 LIGHT CHAIN.  
FT DOMAIN 1550 1675 TRIMERIZATION.

SQ SEQUENCE 1675 AA; 191587 MW; 6C4F2D54801579E2 CRC64;  
Query Match 48.9%; Score 45; DB 1; Length 1675;  
Best Local Similarity 53.3%; Pred. No. 17;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 MPAHLQDDVSFPAP 15  
| | | : | | | : |  
Db 99 MKAHTMTDDVTFWKW 113

## RESULT 6

CLH\_RAT STANDARD; PRT; 1675 AA.  
AC P11442;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Clathrin heavy chain.  
GN CLTC.

OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=88097376; PubMed=3480512;  
RA Kirchhausen T., Harrison S.C., Chow E.P., Mattaliano R.J.,  
RA Ramchandran K.L., Smart J., Brosius J.;  
RT "Clathrin heavy chain: molecular cloning and complete primary  
RT structure.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:8805-8809(1987).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-493.

CC MEDLINE=99043510; PubMed=9827808;  
CC Ter Haar E., Musacchio A., Harrison S.C., Kirchhausen T.;  
CC "Atomic structure of clathrin: a beta propeller terminal domain joins  
CC an alpha zigzag linker.";  
CC Cell 95:563-573(1998).  
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF  
CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES  
CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE  
CC TRANS GOLGI NETWORK.

CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3  
CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE  
CC PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE  
CC PH AND THE CONCENTRATION OF CALCIUM.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND  
CC VESICLES.

CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; J03583; AAA40874.1; -  
DR PIR; A39941; LRTH  
DR PIR; A39941; LBPO; 06-APR-99.  
DR InterPro; IPR001473; Clathrin\_propel.  
DR InterPro; IPR000547; Clathrin\_repeat.  
DR Pfam; PF00637; Clathrin; 7.  
DR Pfam; PF01394; Clathrin\_propel; 7.  
DR SMART; SM00299; CLH; 7.  
KW Coated pits; 3D-structure.

FT DOMAIN 1 479 GLOBULAR TERMINAL DOMAIN.  
FT DOMAIN 480 523 FLEXIBLE LINKER.  
FT DOMAIN 524 1675 HEAVY CHAIN ARM.  
FT DOMAIN 524 634 DISTAL SEGMENT.  
FT DOMAIN 639 1675 PROXIMAL SEGMENT.  
FT DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,  
FT INVOLVED IN LATTICE DISASSEMBLY  
FT (POTENTIAL).  
FT BINDING 1213 1522 LIGHT CHAIN (BY SIMILARITY).

FT DOMAIN 1550 1675 TRIMERIZATION (BY SIMILARITY).  
 SQ SEQUENCE 1675 AA; 191598 MW; C10F54C7ED8C5A61 CRC64;  
 Query Match 48.9%; Score 45; DB 1; Length 1675;  
 Best Local Similarity 53.3%; Pred. No. 17;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MPALLQDDVSPAW 15  
 { | | : | | : | |  
 Db 99 MKAHTMTDDVTWKW 113

RESULT 7  
 VIT2\_FUNHE STANDARD; PRT; 1687 AA.  
 AC Q98993;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Vitellinogen II precursor (VTG II) [Contains: Lipovitellin 1 (LV1);  
 DE Phosvitin (PV); Lipovitellin 2 (LV2); YP 69].  
 OS Fundulus heteroclitus (Killifish) (Mummichog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;  
 OC Cyprinodontiformes; Fundulidae; Fundulus.  
 OX NCBI\_TaxID=8078;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-35.  
 RC TISSUE=Liver;  
 RA Lafleur G.J. Jr., Byrne B.M., Haux C., Greenberg R.M., Wallace R.A.;  
 RA "Liver-derived cDNAs: vitellinogens and vitelline envelope protein  
 precursors (choriogenins).";  
 RL Int. Symp. Reprod. Physiol. Fish 5:336-338(1995).  
 CC -!- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE  
 CC SOURCES OF NUTRIENTS DURING EARLY DEVELOPMENT OF OVIPAROUS  
 CC ORGANISMS.  
 CC -!- TISSUE SPECIFICITY: PRODUCED BY THE LIVER, SECRETED INTO THE BLOOD  
 CC AND THEN SEQUESTERED BY RECEPTOR MEDIATED ENDOCYTOSIS INTO GROWING  
 CC OOCYTES, WHERE IT IS GENERALLY CLEAVED, GIVING RISE TO THE  
 CC RESPECTIVE YOLK COMPONENTS LIPOVITELLINS AND PHOSVITIN.  
 CC -!- INDUCTION: BY STEROIDS (ESTROGEN). EXPRESSION OF VTG II IS LOWER  
 CC THAN THAT OF VTG I.  
 CC -!- P.M.: PHOSVITIN, AN EGG YOLK STORAGE PROTEIN, IS ONE OF THE MOST  
 CC HIGHLY PHOSPHORYLATED (10+) PROTEINS IN NATURE (BY SIMILARITY).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U70826; AAB17152.1; -;  
 CC InterPro: IPR001747; Lipid transprt\_N.  
 CC InterPro: IPR001846; VWF\_D.  
 CC Pfam: PF00094; vwd; 1.  
 CC Pfam: PF01347; Vitellogenin\_N; 1.  
 CC SMART: SM00216; VWD; 1.  
 CC Glycoprotein; Phosphorylation; Storage protein; Signal.  
 CC SIGNAL 1 15 POTENTIAL.  
 CC CHAIN 16 1687 VITELLOGENIN II.  
 CC CHAIN 16 ? LIPOVITELLIN 1.  
 CC CHAIN ? ? PHOSVITIN.  
 CC CHAIN ? 1687 LIPOVITELLIN 2.  
 CC DOMAIN 1059 1062 POLY-GLU.  
 CC DOMAIN 1088 1169 SER-RICH.  
 CC FT CARBOHYD 941 941 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 954 954 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 1004 1004 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 1019 1019 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1083 1083 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1142 1142 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1179 1179 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1257 1257 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1292 1292 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1342 1342 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1361 1361 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1366 1366 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1390 1390 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1577 1577 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1655 1655 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1687 AA; 186005 MW; 4965BB9DBF84928F CRC64;  
 Query Match 47.8%; Score 44; DB 1; Length 1687;  
 Best Local Similarity 61.5%; Pred. No. 26;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PAHLQDDVSEFA 14  
 | | | | : | : |  
 Db 831 PAHLKSDISMKA 843

RESULT 8  
 LOX2\_PIG STANDARD; PRT; 662 AA.  
 ID LOX2\_PIG  
 AC P16469;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Arachidonate 12-lipoxygenase, 12S-type (EC 1.13.11.31) (12-LOX).  
 GN ALOX12.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Leukocyte.  
 RX MEDLINE=90192763; PubMed=2315307;  
 RA Yoshimoto T., Suzuki H., Yamamoto S., Takai T., Yokoyama C.,  
 RA Tanabe T.;  
 RT "Cloning and sequence analysis of the cDNA for arachidonate 12-  
 RT lipoxygenase of porcine leukocytes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2142-2146(1990).  
 RN [2]  
 RP MUTAGENESIS.  
 RX MEDLINE=94137772; PubMed=8305485;  
 RA Suzuki H., Kishimoto K., Yoshimoto T., Yamamoto S., Kanai F.,  
 RA Ebina Y., Miyatake A., Tanabe T.;  
 RT "Site-directed mutagenesis studies on the iron-binding domain and the  
 RT determinant for the substrate oxygenation site of porcine leukocyte  
 RT arachidonate 12-lipoxygenase.";  
 RL Biochim. Biophys. Acta 1210:308-316(1994).  
 CC -!- FUNCTION: OXYGENASE AND 14,15-LEUKOTRIENE A4 SYNTHASE ACTIVITY.  
 CC -!- CATALYTIC ACTIVITY: Arachidonate + O(2) = (5Z,8Z,10E,14Z)-(12S)-  
 CC 12-hydroperoxyicoso-5,8,10,14-tetraenoate.  
 CC -!- COFACTOR: IRON (CONTAINS 0.45 ATOM OF IRON PER MOLECULE).  
 CC -!- PATHWAY: Leukotrienes biosynthesis.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- TISSUE SPECIFICITY: LEUKOCYTES, PITUITARY GLAND, LUNG AND IN  
 CC VERY SMALL AMOUNT IN JEJUNUM AND SPLEEN.  
 CC -!- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M31417; AAA31068.1; -;  
 CC EMBL; D10621; BAA01471.1; -;

```

DR EMBL; D10616; BAA01471.1; JOINED.
DR EMBL; D10617; BAA01471.1; JOINED.
DR EMBL; D10618; BAA01471.1; JOINED.
DR EMBL; D10619; BAA01471.1; JOINED.
DR EMBL; D10620; BAA01471.1; JOINED.
DR F1R; A35087; A35087.
DR HSP; P12530; ILOX.
DR InterPro; IPR000907; Lipoxigenase.
DR InterPro; IPR001024; Lipoxigenase_LH2.
DR Pfam; PF00305; lipoxigenase; 1.
DR Pfam; PF01477; PLAT; 1.
DR PRINTS; PR00087; LIPOXYGENASE.
DR SMART; SM00308; LH2; 1.
DR PROSITE; PS00081; LIPOXYGENASE.2; 1.
DR PROSITE; PS00711; LIPOXYGENASE_L; 1.
KW Oxidoreductase; Dioxigenase; Iron; Leukotriene biosynthesis.
FT INIT_MET 0
FT METAL 360 360 IRON (BY SIMILARITY).
FT METAL 365 365 IRON (BY SIMILARITY).
FT METAL 540 540 IRON (BY SIMILARITY).
FT METAL 662 662 IRON (BY SIMILARITY).
SQ SEQUENCE 662 AA; 74912 MW; 71481B20E368183 CRC64;

Query Match 45.7%; Score 42; DB 1; Length 662;
Best Local Similarity 66.7%; Pred. NO. 20;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 HLLQDDVSFPW 15
DB 68 HLLQDDAWFCNW 79

RESULT 9
GLN1_MAIZE
ID GLN1_MAIZE STANDARD; PRT; 357 AA.
AC P38559;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutamine synthetase root isozyme 1 (EC 6.3.1.2) (Glutamate--ammonia
DE ligase) (GSI22).
GN GLN6 OR GSI-1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. A188; TISSUE=Seedling;
RX MEDLINE=94033318; PubMed=8106013;
RA Li M.-G., Villemur R., Hussey P.J., Silflow C.D., Gantt J.S.,
RA Snustad D.P.;
RT "Differential expression of six glutamine synthetase genes in Zea
RT mays.;"
RL Plant Mol. Biol. 23:401-407(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Golden cross Bantam T51; TISSUE=Leaf;
RA Sakakibara H., Kawabata S., Takahashi H., Hase T., Sugiyama T.;
RT "Molecular cloning of the family of glutamine synthetase genes from
RT maize: expression of genes for glutamine synthetase and ferredoxin-
RT dependent glutamate synthase in photosynthetic and non-photosynthetic
RT tissues.;"
RL Plant Cell Physiol. 33:49-58(1992).
RN [3]
RP REVISIONS.
RA Sakakibara H.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PLAYS A ROLE IN THE FLOW OF NITROGEN INTO NITROGENOUS
CC ORGANIC COMPOUNDS.
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) -> ADP + phosphate +
CC L-glutamine.

```

```

CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: FOUND MAINLY IN THE CORTICAL TISSUES OF
CC SEEDLING ROOTS, AND IN THE ROOT TIP.
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X65926; CAA46719.1; -.
DR EMBL; D14579; BAA03433.1; -.
DR F1R; S39477; S39477.
DR MaizeDB; 17151; -.
DR InterPro; IPR001691; GLN synth.
DR InterPro; IPR001637; GLNAdenyltn.
DR Pfam; PF00120; gln-synt; 1.
DR PROSITE; PS00180; GLNA_1; 1.
DR PROSITE; PS00181; GLNA_ATP; 1.
KW Ligase; Multigene family.
FT CONFLICT 48
FT CONFLICT 48
SQ SEQUENCE 357 AA; 39250 MW; 912A5E3BAF9CC2B8 CRC64;

Query Match 45.1%; Score 41.5; DB 1; Length 357;
Best Local Similarity 69.2%; Pred. NO. 12;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 5 LLLQDDVSFP-AMP 16
DB 134 LLLQDDVSFPGLWP 146

RESULT 10
KPC2_CAEEL
ID KPC2_CAEEL STANDARD; PRT; 707 AA.
AC P34885;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Protein kinase C-like 2 (EC 2.7.1.-) (PKC1B).
GN KIN-13 OR PKC1B.
OS Caenorhabditis elegans
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94179345; PubMed=8132661;
RA Land M., Islas-Trejo A., Freedman J.H., Rubin C.S.;
RT "Structure and expression of a novel, neuronal protein kinase C
RT (PKC1B) from Caenorhabditis elegans. PKC1B is expressed selectively
RT in neurons that receive, transmit, and process environmental
RT signals.;"
RL J. Biol. Chem. 269:9234-9244(1994).
RN [2]
RP FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH MEMBRANES AND THE
CC CYTOSKELETON.
CC -!- TISSUE SPECIFICITY: EXPRESSED SELECTIVELY IN NEURONS THAT RECEIVE,
CC TRANSMIT AND PROCESS ENVIRONMENTAL SIGNALS.
CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC PKC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U00181; AAA18259.1; -;  
 DR EMBL; U00965; AAA17996.1; -;  
 DR PIR; A53530; A53530.  
 DR HSP; P28867; IPTQ.  
 DR InterPro; IPR000008; C2.  
 DR InterPro; IPR002219; DAG\_PE-bind.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR000961; pkinase\_C.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00130; DAG\_PE-bind; 2.  
 DR Pfam; PF00168; C2; 1.  
 DR Pfam; PF00433; pkinase\_C; 1.  
 DR PRINTS; PR000008; DAGPEDOMAIN.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00109; C1; 2.  
 DR SMART; SM00239; C2; 1.  
 DR SMART; SM00133; S\_TK\_X; 1.  
 DR SMART; SM00220; S\_TK; 1.  
 DR PROSITE; PS00004; C2\_DOMAIN\_2; 1.  
 DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
 DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 2.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding; Zinc;  
 KW Phorbol-ester binding; Repeat; Membrane; Cytoskeleton.  
 FT DOMAIN 1 98  
 FT DOMAIN 171 221  
 FT DOMAIN 249 298  
 FT DOMAIN 378 638  
 FT NP\_BIND 384 392  
 FT BINDING 407 407  
 FT ACT\_SITE 502 502  
 FT ACT\_SITE 502 502  
 SQ SEQUENCE 707 AA; 3DC762C8A7A7BA64 CRC64;

Query Match 44.6%; Score 41; DB 1; Length 707;  
 Best Local Similarity 54.5%; Pred. No. 32;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLOQDVSPAP 15  
 Db 589 ILNDDVLYPVW 599

RESULT 11

KPCE RABIT  
 ID KPCE\_RABIT STANDARD; PRT; 736 AA.  
 AC P10830;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Protein kinase C, epsilon type (EC 2.7.1.-) (nPKC-epsilon).  
 GN PKCE.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8823367; PubMed=3370672;  
 RA Omo S., Akita Y., Konno Y., Imajoh S., Suzuki K.;  
 RT "A novel phorbol ester receptor/protein kinase, nPKC, distantly  
 RT related to the protein kinase C family";  
 RL Cell 53:731-741(1988).  
 CC -!- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,

CC SERINE- AND THREONINE-SPECIFIC ENZYME.  
 CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN  
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS  
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.  
 CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 CC BINDING DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC PKC SUBFAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M20014; AAA31426.1; -;  
 DR PIR; A29880; KIRBCE.  
 DR HSP; P28867; IPTQ.  
 DR InterPro; IPR000008; C2.  
 DR InterPro; IPR002219; DAG\_PE-bind.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR000961; pkinase\_C.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00130; DAG\_PE-bind; 2.  
 DR Pfam; PF00168; C2; 1.  
 DR Pfam; PF00433; pkinase\_C; 1.  
 DR PRINTS; PR000008; DAGPEDOMAIN.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00109; C1; 2.  
 DR SMART; SM00239; C2; 1.  
 DR SMART; SM00133; S\_TK\_X; 1.  
 DR SMART; SM00220; S\_TK; 1.  
 DR PROSITE; PS00004; C2\_DOMAIN\_2; 1.  
 DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 1.  
 DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 2.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Repeat; ATP-binding; Transferase; Phosphorylation;  
 KW Serine/threonine-protein kinase; Phorbol-ester binding; Zinc.  
 FT DOMAIN 1 99  
 FT DOMAIN 170 220  
 FT DOMAIN 243 292  
 FT DOMAIN 407 667  
 FT NP\_BIND 413 421  
 FT BINDING 436 436  
 FT ACT\_SITE 531 531  
 FT MOD\_RES 702 702  
 FT MOD\_RES 709 709  
 SQ SEQUENCE 736 AA; 83515 MW; 261C4FE59E9BEE CRC64;

Query Match 44.6%; Score 41; DB 1; Length 736;  
 Best Local Similarity 54.5%; Pred. No. 33;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLOQDVSPAP 15  
 Db 618 ILNDDVLYPVW 628

RESULT 12

KPCE HUMAN  
 ID KPCE\_HUMAN STANDARD; PRT; 737 AA.  
 AC Q02156; Q9UE81;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein kinase C, epsilon type (EC 2.7.1.-) (nPKC-epsilon).  
 GN PKCE OR PKCE.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93003318; PubMed=1382605;  
 RA Basta P., Strickland M.B., Holmes W., Loomis C.R., Ballas L.M.,  
 RA Burds D.J.;  
 RT "Sequence and expression of human protein kinase C-epsilon.";  
 RL Biochim. Biophys. Acta 1132:154-160(1992).  
 RN [2]  
 RP SEQUENCE OF 1-116 FROM N.A.  
 RA Waterston R.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,  
 CC SERINE- AND THREONINE-SPECIFIC ENZYME.  
 CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN  
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS  
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.  
 CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 CC BINDING DOMAINS  
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC PKC SUBFAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; X65293; CAA46388.1; -;  
 CC EMBL; U51244; AAD08855.1; -;  
 CC PIR; S28942; S28942.  
 CC DR HSP; P28867; IPTQ.  
 CC DR Genew; HGNC:9401; PKCE.  
 CC DR MIM; 176975; -;  
 CC DR InterPro; IPR000008; C2.  
 CC DR InterPro; IPR002219; DAG\_PE-bind.  
 CC DR InterPro; IPR000713; Euk\_pkinase.  
 CC DR InterPro; IPR000961; Pkinase\_C.  
 CC DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 CC DR Pfam; PF00069; pkinase; 1.  
 CC DR Pfam; PF00130; DAG\_PE-bind; 2.  
 CC DR Pfam; PF00168; C2; 1.  
 CC DR Pfam; PF00433; pkinase\_C; 1.  
 CC DR PRINTS; PR00008; DAGPEDOMAIN.  
 CC DR ProDom; PD000001; Euk\_pkinase; 1.  
 CC DR SMART; SM00109; C1; 2.  
 CC DR SMART; SM00239; C2; 1.  
 CC DR SMART; SM00133; S-TK.X; 1.  
 CC DR SMART; SM00220; S-TKC; 1.  
 CC DR PROSITE; PS00004; C2\_DOMAIN\_2; 1.  
 CC DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
 CC DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 2.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 CC DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 CC Repeat; ATP-binding; Transferase; Phosphorylation;  
 CC Serine/threonine-protein kinase; Phorbol-ester binding; Zinc.  
 CC DOMAIN 1 99 C2 DOMAIN.  
 CC FT DOMAIN 170 220 PHORBOL-ESTER AND DAG BINDING 1.  
 CC FT DOMAIN 243 292 PHORBOL-ESTER AND DAG BINDING 2.  
 CC FT DOMAIN 408 668 PROTEIN KINASE.  
 CC FT NP\_BIND 414 422 ATP (BY SIMILARITY).  
 CC FT BINDING 437 437 ATP (BY SIMILARITY).  
 CC FT ACT\_SITE 532 532 BY SIMILARITY.  
 CC FT MOD\_RES 703 703 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 CC FT MOD\_RES 710 710 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 CC SEQUENCE 737 AA; 83673 MW; 85032D0A091A1F7F CRC64;

Query Match 44.6%; Score 41; DB 1; Length 737;  
 Best Local Similarity 54.5%; Pred. No. 33;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 LIQDDVSFFAW 15  
 Db 619 ILHDDVLYPFW 629  
 RESULT 13  
 KCPE\_MOUSE STANDARD; PRT; 737 AA.  
 ID KCPE\_MOUSE  
 AC P16054;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protein Kinase C, epsilon type (EC 2.7.1.-) (nPKC-epsilon).  
 GN PKCE OR PKCE OR PKCEA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89137541; PubMed=2917656;  
 RA Schaap D., Parker P.J., Bristol A., Kriz R., Knopf J.;  
 RT "Unique substrate specificity and regulatory properties of  
 RT PKC-epsilon: a rationale for diversity.";  
 RL FEBS Lett. 243:351-357(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Brain;  
 RX MEDLINE=98127436; PubMed=9467942;  
 RA Wang Q.J., ACS P., Goodnight J., Blumberg P.M., Mischak H.,  
 RA Mushinski J.F.;  
 RT "The catalytic domain of PKC-epsilon, in reciprocal PKC-delta and -  
 RT epsilon chimeras, is responsible for conferring tumorigenicity to  
 RT NIH3T3 cells, whereas both regulatory and catalytic domains of  
 RT PKC-epsilon contribute to in vitro transformation.";  
 RL Oncogene 16:53-60(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Wheeler D.L.;  
 RA Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,  
 CC SERINE- AND THREONINE-SPECIFIC ENZYME.  
 CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN  
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS  
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.  
 CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 CC BINDING DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC PKC SUBFAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; AF028009; AAB84189.1; -;  
 CC EMBL; AF325507; AAG53692.1; -;  
 CC PIR; S02270; KIMSCF.  
 CC HSSP; P28867; IPTQ.  
 CC MGL; MGI:97599; Pkce.  
 CC InterPro; IPR000008; C2.  
 CC InterPro; IPR002219; DAG\_PE-bind.  
 CC InterPro; IPR000713; Euk\_pkinase.

```

DR InterPro; IPR000961; Pkinase.C.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00130; DAG_PE-bind; 2.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00433; pkinase.C; 1.
DR PRINTS; PR00008; DAGPEPDOMAIN.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00109; C1; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00133; S_TK_X; 1.
DR SMART; SM00220; S_TK_X; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Repeat; ATP-binding; Transferase; Phosphorylation;
KW Serine/threonine-protein kinase; Phorbol-ester binding; Zinc.
FT DOMAIN 1 99
FT DOMAIN 170 220 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 243 292 PHORBOL-ESTER AND DAG BINDING 2.
FT NP_BIND 414 422 ATP (BY SIMILARITY).
FT BINDING 437 437 ATP (BY SIMILARITY).
FT ACT_SITE 532 532 BY SIMILARITY.
FT MOD_RES 703 703 PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT MOD_RES 710 710 PHOSPHORYLATION (AUTO-) (POTENTIAL).
SQ SEQUENCE 737 AA; 83560 MW; 7AEBB8CC10C99F57 CRC64;

Query Match 44.6%; Score 41; DB 1; Length 737;
Best Local Similarity 54.5%; Pred. No. 33;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLQDDVSFPAP 15
Db 619 ILHDDVLYPVW 629
:| ||| :|

RESULT 14
KPC2_RAT
ID KPC2_RAT STANDARD; PRT; 737 AA.
AC P09216;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Protein kinase C, epsilon type (EC 2.7.1.-) (nPKC-epsilon).
GN PKCE OR PKCE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP TISSUE=Brain;
RC SEQUENCE FROM N.A.
RX MEDLINE=88198270; PubMed=2834397;
RA Ono Y., Fujii T., Ogita K., Kikkawa U., Igarashi K., Nishizuka Y.;
RT "The structure, expression, and properties of additional members of
RT the protein kinase C family.";
RL J. Biol. Chem. 263:6927-6932(1988).
RN [2]
RP SEQUENCE OF 135-297 FROM N.A.
RX MEDLINE=88083621; PubMed=3691811;
RA Ono Y., Fujii T., Ogita K., Kikkawa U., Igarashi K., Nishizuka Y.;
RT "Identification of three additional members of rat protein kinase C
RT family: delta-, epsilon- and zeta-subspecies.";
RL FEBS Lett. 226:125-128(1987).
CC -!- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,
CC SERINE- AND THREONINE-SPECIFIC ENZYME.
CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.

```

-!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG BINDING DOMAINS.  
 -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. PKC SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M18331; AAA1872.1; -  
 PIR; B28163; KIRCE.  
 HSSP; P28867; IPTQ.  
 InterPro; IPR000008; C2.  
 InterPro; IPR002219; DAG\_PE-bind.  
 InterPro; IPR000719; Euk\_pkinase.  
 InterPro; IPR000961; Pkinase.C.  
 InterPro; IPR002290; Ser\_thr\_pkinase.  
 Pfam; PF00069; Pkinase; 1.  
 Pfam; PF00130; DAG\_PE-bind; 2.  
 Pfam; PF00168; C2; 1.  
 Pfam; PF00433; pkinase.C; 1.  
 PRINTS; PR00008; DAGPEPDOMAIN.  
 ProDom; PD000001; Euk\_pkinase; 1.  
 SMART; SM00109; C1; 2.  
 SMART; SM00239; C2; 1.  
 SMART; SM00133; S\_TK\_X; 1.  
 SMART; SM00220; S\_TK\_X; 1.  
 PROSITE; PS00004; C2\_DOMAIN\_2; 1.  
 PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 2.  
 PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 2.  
 PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 Repeat; ATP-binding; Transferase; Phosphorylation;  
 Serine/threonine-protein kinase; Phorbol-ester binding; Zinc.  
 FT DOMAIN 1 99  
 FT DOMAIN 170 220 PHORBOL-ESTER AND DAG BINDING 1.  
 FT DOMAIN 243 292 PHORBOL-ESTER AND DAG BINDING 2.  
 FT NP\_BIND 414 422 ATP (BY SIMILARITY).  
 FT BINDING 437 437 ATP (BY SIMILARITY).  
 FT ACT\_SITE 532 532 BY SIMILARITY.  
 FT MOD\_RES 703 703 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 FT MOD\_RES 710 710 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
 SQ SEQUENCE 737 AA; 83478 MW; 6AD6999FEFDD2659F CRC64;

Query Match 44.6%; Score 41; DB 1; Length 737;  
 Best Local Similarity 54.5%; Pred. No. 33;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLQDDVSFPAP 15  
 Db 619 ILHDDVLYPVW 629  
 :| ||| :|

RESULT 15  
 KPC2\_APLCA  
 ID KPC2\_APLCA STANDARD; PRT; 743 AA.  
 AC Q16975;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Calcium-independent protein kinase C (EC 2.7.1.-) (APL II).  
 GN PRKC2.  
 OS Aplysia californica (California sea hare).  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspiidea;  
 OC Aplysiidae; Aplysia.  
 OX NCBI\_TaxID=6500;

```

RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE-91332620; PubMed-1869917;
RA  Kruger K.E., Sossin W.S., Sacktor T.C., Bergold P.J., Beushausen S.,
RA  Schwartz J.H.;
RT  "Cloning and characterization of Ca(2+)-dependent and Ca(2+)-
RT  independent PKCs expressed in Aplysia sensory cells.";
RL  J. Neurosci. 11:2303-2313(1991).
RN  [2]
RP  CHARACTERIZATION.
RX  MEDLINE-93194877; PubMed-8449941;
RA  Sossin W.S., Diaz-Arriastia R., Schwartz J.H.;
RT  "Characterization of two isoforms of protein kinase C in the nervous
RT  system of Aplysia californica.";
RL  J. Biol. Chem. 268:5763-5768(1993).
RN  [3]
RP  DOMAIN C2.
RX  MEDLINE-93348616; PubMed-8346555;
RA  Sossin W.S., Schwartz J.H.;
RT  "Ca(2+)-independent protein kinase Cs contain an amino-terminal domain
RT  similar to the C2 consensus sequence.";
RL  Trends Biochem. Sci. 18:207-208(1993).
RN  [4]
RP  CHARACTERIZATION.
RX  MEDLINE-98334636; PubMed-9668085;
RA  Pepio A.M., Fan X., Sossin W.S.;
RT  "The role of C2 domains in Ca2+-activated and Ca2+-independent protein
RT  kinase Cs in Aplysia.";
RL  J. Biol. Chem. 273:19040-19048(1998).
RN  [5]
RP  ERRATUM.
RA  Pepio A.M., Fan X., Sossin W.S.;
RL  J. Biol. Chem. 273:22856-22856(1998).
RN  [6]
RP  CHARACTERIZATION.
RX  MEDLINE-98138438; PubMed-9477951;
RA  Pepio A.M., Sossin W.S.;
RT  "The C2 domain of the Ca(2+)-independent protein kinase C Apl II
RT  inhibits phorbol ester binding to the C1 domain in a phosphatidic
RT  acid-sensitive manner.";
RL  Biochemistry 37:1256-1263(1998).
CC  -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,
CC  SERINE- AND THREONINE-SPECIFIC ENZYME.
CC  -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC  PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC  THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS (BY
CC  SIMILARITY).
CC  -1- ENZYME REGULATION: REQUIRES HIGH LEVELS OF PHOSPHATIDYL SERINE TO
CC  BE ACTIVATED. THE PRESENCE OF THE C2 DOMAIN LOWERS THE AFFINITY OF
CC  PROTEIN KINASE C ACTIVATORS FOR THE C1 DOMAINS AND THIS INHIBITION
CC  CAN BE REMOVED BY PHOSPHATIDYL SERINE. PHOSPHATIDIC ACID, HOWEVER,
CC  IS MUCH MORE POTENT THAN PHOSPHATIDYL SERINE IN REDUCING C2 DOMAIN-
CC  MEDIATED INHIBITION, SUGGESTING THAT PHOSPHATIDYL SERINE MAY BE A
CC  REQUIRED COFACTOR FOR THE ACTIVATION OF APL II.
CC  -1- SUBCELLULAR LOCATION: Membrane-associated.
CC  -1- TISSUE SPECIFICITY: EXPRESSED IN NERVOUS TISSUES, OVOTESTIS AND
CC  GUT.
CC  -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC  BINDING DOMAINS.
CC  -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC  -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC  PKC SUBFAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL: M94884; AAA27771.1;
DR  HSP: P28867; LPTQ.

```

```

DR  InterPro: IPR000008; C2.
DR  InterPro: IPR002219; DAG_PE-bind.
DR  InterPro: IPR000719; Euk_pkinase.
DR  InterPro: IPR000961; Pkinase_C.
DR  InterPro: IPR002290; Ser_thr_pkinase.
DR  Pfam: PF00130; DAG_PE-bind; 1.
DR  Pfam: PF00168; C2; 1.
DR  Pfam: PF00433; pkinase_C; 1.
DR  PRINTS: PR00008; DAGPEDOMAIN.
DR  ProDom: PD000001; Euk_pkinase; 1.
DR  SMART: SM00109; C1; 2.
DR  SMART: SM00239; C2; 1.
DR  SMART: SM00133; S_TK_X; 1.
DR  SMART: SM00220; S_TKC; 1.
DR  PROSITE: PS00499; C2_DOMAIN_1; FALSE_NEG.
DR  PROSITE: PS00004; C2_DOMAIN_2; 1.
DR  PROSITE: PS00479; DAG_PE_BIND_DOM_1; 2.
DR  PROSITE: PS00081; DAG_PE_BIND_DOM_2; 2.
DR  PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR  PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR  PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW  ATP-binding; Transferase; Serine/threonine-protein kinase;
KW  Phorbol-ester binding; Zinc; Repeat.
FT  DOMAIN 1 108 C2 DOMAIN.
FT  DOMAIN 177 226 PHORBOL-ESTER AND DAG BINDING 1.
FT  DOMAIN 248 297 PHORBOL-ESTER AND DAG BINDING 2.
FT  DOMAIN 403 663 PROTEIN KINASE.
FT  NP_BIND 409 417 ATP (BY SIMILARITY).
FT  BINDING 432 527 ATP (BY SIMILARITY).
FT  ACT_SITE 527 527 BY SIMILARITY.
SQ  SEQUENCE 743 AA; 84413 MW; 4C982C563CA2B659 CRC64;

Query Match 44.6%; Score 41; DB 1; Length 743;
Best Local Similarity 54.5%; Pred. No. 34;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLQDDVSFPAP 15
Db 614 ILHDDVLYPVW 624

Search completed: May 14, 2003, 15:31:32
Job time : 17 secs

```



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:22:26 ; Search time 29 Seconds  
(without alignments)  
113.681 Million cell updates/sec

Title: US-10-016-725-16

Perfect score: 92

Sequence: 1 MPAHLQDDVSPFAMP 16

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mic:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Summary			Description
	Score	Query Match	Length DB ID	
1	49	53.3	711 2	Q9AI02
2	48	52.2	359 6	Q95MI7
3	48	52.2	359 6	Q9BG81
4	48	52.2	359 6	Q8SQ76
5	47	51.1	688 16	Q984P3
6	45	48.9	180 13	Q8UUR0
7	45	48.9	1675 13	Q8UUR1
8	44	47.8	68 5	Q9BPJ7
9	44	47.8	506 5	Q9GR90
10	43.5	47.3	518 10	Q81301
11	43	46.7	136 2	Q49796
12	43	46.7	208 4	Q9H718
13	43	46.7	437 16	Q99Z61
14	43	46.7	674 16	Q92QX2
15	43	46.7	753 10	Q8S864
16	43	46.7	795 10	Q9LXX0

## SUMMARIES

17	43	46.7	800	10	082312
18	43	46.7	1725	13	Q8UW61
19	43	46.7	1793	5	Q9NKKJ4
20	42.5	46.2	276	16	Q9KYM9
21	42	45.7	79	4	Q12938
22	42	45.7	255	10	Q8VYH8
23	42	45.7	270	16	Q9KY29
24	42	45.7	551	10	Q94FS6
25	42	45.7	847	16	Q8YOC4
26	42	45.7	1280	10	Q9LFP9
27	42	45.7	2380	16	Q8XVE8
28	41.5	45.1	388	2	Q9WVW9
29	41.5	45.1	627	16	Q9PB34
30	41.5	45.1	1035	10	Q8S389
31	41	44.6	160	16	Q9K6U2
32	41	44.6	227	16	Q33336
33	41	44.6	240	11	Q63432
34	41	44.6	338	5	Q9XV78
35	41	44.6	357	13	Q9YGM2
36	41	44.6	387	16	Q9PEV4
37	41	44.6	490	13	Q42492
38	41	44.6	567	4	Q8TEN4
39	41	44.6	660	16	Q8ZP78
40	41	44.6	660	16	Q8Z795
41	41	44.6	681	16	Q8UGA9
42	41	44.6	685	16	Q8YGA9
43	41	44.6	707	5	Q20953
44	41	44.6	772	6	Q95JG9
45	41	44.6	847	16	Q8ZQ26

## ALIGNMENTS

## RESULT 1

Q9AI02 ID Q9AI02 PRELIMINARY; PRT; 711 AA.  
AC Q9AI02;  
DT 01-JUN-2001 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)  
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)  
DE NGRD (Fragment).  
GN NGRD.  
OS Photorhabdus luminescens (Xenorhabdus luminescens).  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Photorhabdus.  
OX NCBI\_taxID=29488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NC19;  
RX MEDLINE=21225535; PubMed=11325940;  
RA Cliche T.A., Bintrim S.B., Horswill A.R., Ebsign J.C.;  
RT "A Phosphopantetheinyl Transferase Homolog Is Essential for  
RT Photorhabdus luminescens To Support Growth and Reproduction of the  
RT Entomopathogenic Nematode Heterorhabditis bacteriophora.";  
RL J. Bacteriol. 183:3117-3126(2001).  
DR EMBL; AF288082; AAKI6083.1;  
DR InterPro; IPR003838; DUF214.  
DR Pfam; PF02687; DUF214; 1.  
ET NON\_TER 711 711  
SQ SEQUENCE 711 AA; 79781 MW; 523ESCB4546CD805 CRC64;

Query Match 53.3%; Score 49; DB 2; Length 711;  
Best Local Similarity 46.7%; Pred.No. 6.8;  
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAHLQDDVSPFAMP 16

Db 364 PSRLVLDIIAIPW 378

## RESULT 2

Q95MI7

```

ID Q95MI7 PRELIMINARY; PRT; 359 AA.
AC Q95MI7:
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Stearoyl coenzyme A desaturase (EC 1.14.99.5).
GN SCD.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
FN SEQUENCE FROM N.A.
RP Yalyaoui M.H., Sanchez A., Folch J.M.;
RA "Nucleotide sequence of the goat stearoyl coenzyme A desaturase cDNA
RT and gene structure.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP Yalyaoui M.H., Sanchez A., Folch J.M.;
RA "Partial nucleotide sequence of the goat stearoyl coenzyme A
RT desaturase cDNA and gene structure.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF33909; AAK61862.1; -
DR EMBL; AF422171; AAL29305.1; -
DR EMBL; AF422166; AAL29305.1; JOINED.
DR EMBL; AF422167; AAL29305.1; JOINED.
DR EMBL; AF422168; AAL29305.1; JOINED.
DR EMBL; AF422169; AAL29305.1; JOINED.
DR EMBL; AF422170; AAL29305.1; JOINED.
DR InterPro; IPR001522; Desaturase.
DR InterPro; IPR001225; FA_desaturase.
DR Pfam; PF00487; FA_desaturase; 1.
DR ProDom; PD002221; Desaturase; 1.
DR PROSITE; PS00476; FATTY_ACID_DESATUR_1; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 359 AA; 41586 MW; 052E5BE2B4463F88 CRC64;

Query Match 52.2%; Score 48; DB 6; Length 359;
Best Local Similarity 72.7%; Pred. No. 4.9;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
Db 1 MPAHLQEEIS 11

RESULT 3
ID Q9BG81 PRELIMINARY; PRT; 359 AA.
AC Q9BG81:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Acyl-CoA desaturase (EC 1.14.99.5) (Stearoyl-CoA desaturase) (Fatty
acid desaturase) (Delta(9)-desaturase).
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
FN SEQUENCE FROM N.A.
RP TISSUE-MAMMARY GLAND;
RA Bernard L., Leroux C., Hayes H., Gautier M., Chilliard Y., Martin P.;
RT "Genomic organization, chromosomal localization, and the complete 5 kb
RT cDNA sequence of the caprine SCD gene.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: STEAROYL-COA + AH(2) + O(2) = OLEOYL-COA + A +
CC 2 H(2)O.
CC -1- COFACTOR: IRON (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC

```

```

CC RETICULUM (BY SIMILARITY).
CC -1- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC AND/OR BE INVOLVED IN METAL ION BINDING (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
DR EMBL; AF325499; AAK01666.1; -
DR InterPro; IPR001522; Desaturase.
DR InterPro; IPR001225; FA_desaturase.
DR Pfam; PF00487; FA_desaturase; 1.
DR PRINTS; PR00075; FACDDSATRASE.
DR ProDom; PD002221; Desaturase; 1.
DR PROSITE; PS00476; FATTY_ACID_DESATUR_1; 1.
KW Endoplasmic reticulum; Fatty acid biosynthesis; Iron; Oxidoreductase;
KW Transmembrane.
SQ SEQUENCE 359 AA; 41582 MW; D27E5D55B44630CA CRC64;

Query Match 52.2%; Score 48; DB 6; Length 359;
Best Local Similarity 72.7%; Pred. No. 4.9;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
Db 1 MPAHLQEEIS 11

RESULT 4
ID Q8SQ76 PRELIMINARY; PRT; 359 AA.
AC Q8SQ76:
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Stearoyl-CoA desaturase (EC 1.14.99.5).
GN SCD.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
FN SEQUENCE FROM N.A.
RA Glimm D., Dong F., Kennelly J.;
RT "Bovine stearoyl-CoA desaturase gene structure and large scale SNP
RT analysis.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF481919; AAL99940.1; -
DR EMBL; AF481915; AAL99940.1; JOINED.
DR EMBL; AF481916; AAL99940.1; JOINED.
DR EMBL; AF481917; AAL99940.1; JOINED.
DR EMBL; AF481918; AAL99940.1; JOINED.
KW Oxidoreductase.
SQ SEQUENCE 359 AA; 41733 MW; C7C9F6F52A90DD15 CRC64;

Query Match 52.2%; Score 48; DB 6; Length 359;
Best Local Similarity 72.7%; Pred. No. 4.9;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
Db 1 MPAHLQEEIS 11

RESULT 5
ID Q984P3 PRELIMINARY; PRT; 688 AA.
AC Q984P3:
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Excinuclease ABC subunit C.
GN MLE7912.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.

```

```

OX NCB1_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003012; BAB54270.1; -
DR InterPro: IPR000445; HhH.
DR InterPro: IPR003583; HhH.1.
DR InterPro: IPR001943; UvrB/C.
DR InterPro: IPR004791; UvrC.
DR InterPro: IPR001162; UvrC.C.
DR InterPro: IPR000305; UvrC.N.
DR Pfam: PF01541; Excl_endo_N; 1.
DR Pfam: PF02151; Uvr; 1.
DR ProDom: PD005870; UvrC.C; 1.
DR SMART: SM00465; GYIC; 1.
DR SMART: SM00278; HHH1; 2.
DR TIGRFAMs: TIGR00194; uvrC; 1.
DR Complete proteome.
KW SEQUENCE 688 AA; 76147 MW; B10906D705614646 CRC64;

Query Match 51.1%; Score 47; DB 16; Length 688;
Best Local Similarity 61.5%; Pred. No. 14;
Matches 8; Conservative 3; Mismatches 0; Indels 2; Gaps 1;

QY 6 LQDDV--SPFAMP 16
Db 529 IEDDGGSPFAMP 541
::||: |||||

RESULT 6
Q8UUR0 PRELIMINARY; PRT; 180 AA.
AC Q8UUR0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Clathrin heavy-chain (Fragment).
GN CHC.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCB1_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN;
RA Wetley F.R., Hawkins S.F.C., Stewart A., Luzio J.P., Howard J.C.,
RA Jackson A.P.;
RT "Controlled elimination of clathrin heavy-chain expression in DT40
RT lymphocytes inhibits endocytosis but not lysosome biogenesis."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ429072; CAD2057.1; -
DR InterPro: IPR001473; Clathrin_propel.
DR Pfam: PF01394; Clathrin_propel; 4.
DR SMART: SM00299; CLH; 7.
DR NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 20155 MW; 85BC945FA30198E3 CRC64;

Query Match 48.9%; Score 45; DB 13; Length 180;
Best Local Similarity 53.3%; Pred. No. 7.6;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSPFAMP 15
Db 85 MKAHTMTDDVTFWKW 99
|||: |||||

RESULT 7
Q8UUR1 PRELIMINARY; PRT; 1675 AA.
AC Q8UUR1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Clathrin heavy-chain (Fragment).
GN CHC.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCB1_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN;
RA Wetley F.R., Hawkins S.F.C., Stewart A., Luzio J.P., Howard J.C.,
RA Jackson A.P.;
RT "Controlled elimination of clathrin heavy-chain expression in DT40
RT lymphocytes inhibits endocytosis but not lysosome biogenesis."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ429072; CAD2057.1; -
DR InterPro: IPR001473; Clathrin_propel.
DR Pfam: PF01394; Clathrin_propel; 7.
DR SMART: SM00299; CLH; 7.
DR NON_TER 1
FT NON_TER 1675
SQ SEQUENCE 1675 AA; 191599 MW; 337BE0B5423592DA CRC64;

Query Match 48.9%; Score 45; DB 13; Length 1675;
Best Local Similarity 53.3%; Pred. No. 81;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSPFAMP 15
Db 99 MKAHTMTDDVTFWKW 113
|||: |||||

RESULT 8
Q8BPJ7 PRELIMINARY; PRT; 68 AA.
AC Q8BPJ7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Conotoxin scaffold III/IV.
OS Conus textile (Cloth-of-gold cone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC Neogastropoda; Conoidea; Conidae; Conus.
OX NCB1_TaxID=6494;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21105969; PubMed=11158371;
RA Conticello S.G., Gilad Y., Avidan N., Ben-Asher E., Levy Z.,
RA Fainzilber M.;
RT "Mechanisms for Evolving Hypervariability: The Case of Conopeptides."
RL Mol. Biol. Evol. 18:120-131(2001).
DR EMBL: AF214926; AAG60354.1; -
DR NON_TER 1
FT NON_TER 180
SQ SEQUENCE 68 AA; 7603 MW; 54CA322FFA7CB1B0 CRC64;

Query Match 47.8%; Score 44; DB 5; Length 68;
Best Local Similarity 53.3%; Pred. No. 4;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAHLQDDVSPFAMP 16
Db 31 PAERMQDDISFEQHP 45
|||: |||||

```

## RESULT 9

Q9GR90 PRELIMINARY; PRT; 506 AA.  
 AC Q9GR90;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE Putative mannose-specific lectin.  
 OS Polyandrocampa misakiensis.  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
 OC Stolidobranchia; Styelidae; Polyandrocarpa.  
 OX NCBI\_TaxID=7723;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sagara Y., Fujiwara S., Yubisui T.;  
 RT "cDNA clonings from Polyandrocarpa misakiensis.";  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB053114; BAB20045.1; -  
 DR InterPro; IPR005052; Lectin\_leg.  
 DR Pfam; PF03388; Lectin\_leg-like; 1.  
 KW Lectin.  
 SQ SEQUENCE 506 AA; 57282 MW; 81524C9295B9B199 CRC64;

Query Match 47.8%; Score 44; DB 5; Length 506;  
 Best Local Similarity 50.0%; Pred. No. 34;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 HLLQDDVSFPW 15

DB 40 HLIQEDLTVPFW 51

## RESULT 10

O81301 PRELIMINARY; PRT; 518 AA.  
 AC O81301;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE T14P8.1 protein (Pectinesterase-like protein).  
 GN T14P8.1 OR AF4G02320.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA WASHU;  
 RT "The A. thaliana Genome Sequencing Project.";   
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Kalicki J., Elliott G., Cloud J.;  
 RT "The sequence of A. thaliana T14P8.";   
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Waterston R.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF069298; AAC19272.1; -

DR EMBL; AL161494; CAB80725.1; -  
 DR InterPro; IPR000070; Pectinesterase.  
 DR Pfam; PF01095; Pectinesterase; 1.  
 DR PROSITE; PS00503; PECTINESTERASE\_2; 1.  
 SQ SEQUENCE 518 AA; 57894 MW; BECBA58404C701B7 CRC64;

Query Match 47.3%; Score 43.5; DB 10; Length 518;  
 Best Local Similarity 38.9%; Pred. No. 42;  
 Matches 7; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 MPAHL---LQDDVSFPW 15

DB 171 IFGHPGVKVEDVGFPW 188

## RESULT 11

O49796 PRELIMINARY; PRT; 136 AA.  
 AC O49796;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE B2126.F1\_18.  
 GN MLCB2533.I3C.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R.;  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Hamlin N., Churcher C.M.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93188700; PubMed=8446027;  
 RA Eglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;  
 RT "Use of an ordered cosmid library to deduce the genomic organization  
 of Mycobacterium leprae.";   
 RL Mo., Microbiol. 7:197-206(1993).  
 DR EMBL; U00017; AAA17210.1; -  
 DR EMBL; AL035310; CAA22927.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 136 AA; 14872 MW; F9DC6B023643ED34 CRC64;

Query Match 46.7%; Score 43; DB 2; Length 136;  
 Best Local Similarity 70.0%; Pred. No. 12;  
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAHLQDDVS 11

DB 36 PAHVLDQDLT 45

## RESULT 12

Q9H718 PRELIMINARY; PRT; 208 AA.  
 ID Q9H718  
 AC Q9H718;

```

DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE FLJ00096 protein (Fragment).
GN FLJ00096
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Chara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RL spleen."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK024492; BAB15782.1;
FT NON_TER
SQ SEQUENCE 208 AA; 22275 MW; 05781A54D80692C1 CRC64;

Query Match 46.7%; Score 43; DB 4; Length 208;
Best Local Similarity 46.7%; Pred. No. 20;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFPWP 16
Db 187 PSRLPDASNPWP 201

RESULT 13
ID Q99261 PRELIMINARY; PRT; 437 AA.
AC Q99261;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative chloride channel protein.
GN SPV1379.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Szate S., Suarov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: AE006575; AAK34204.1;
DR InterPro: IPR001807; Cl-channel_volt.
DR Pfam: PF00654; voltage_CIC; 1.
DR PRINTS; PR00762; CLCHANNEL.
KW Complete proteome.
SQ SEQUENCE 437 AA; 47717 MW; 28DA453AC7EBC410 CRC64;

Query Match 46.7%; Score 43; DB 16; Length 437;
Best Local Similarity 54.3%; Pred. No. 43;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLOQDDVSFPWP 15
Db 117 ILOQDDVSFPW 127

RESULT 14
Q920X2 PRELIMINARY; PRT; 674 AA.
ID Q920X2
AC Q920X2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)

```

```

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Probable exinclease ABC subunit C protein.
GN UVRC OR R01171 OR SMC00602.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subphylum; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [-]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Anpe F., Batut J.,
RA Boistard P., Becker A., Bonty M., Cadieu E., Dreano S., Gloux S.,
RA Godelle T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591786; CAC45750.1;
DR InterPro: IPR000445; Hhh.
DR InterPro: IPR003583; Hhh.
DR InterPro: IPR001943; UvrB/C.
DR InterPro: IPR004791; UvrC.
DR InterPro: IPR001162; UvrC.
DR InterPro: IPR000305; UvrC_N.
DR Pfam: PF01541; Exci_endo_N; 1.
DR Pfam: PF02151; UVR; 1.
DR ProDom: PD005870; UvrC_C; 1.
DR SMART: SM00278; Hhh1; 2.
DR TIGRFAMs: TIGR00194; uvrC; 1.
KW Complete proteome.
SQ SEQUENCE 674 AA; 74417 MW; 14FCFE72FB39A580 CRC64;

Query Match 46.7%; Score 43; DB 16; Length 674;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 DDVSFPWP 16
Db 516 EDAGFPWP 524

RESULT 15
Q8S864 PRELIMINARY; PRT; 753 AA.
ID Q8S864
AC Q8S864;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 85.8 kDa protein.
GN OSJNBA0061K21.13.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Spiegel L.A., Nascimiento L.U., de la Bastide M., Vil D.M.,
RA Preston R.R., Huang E.N., See L.H., Shah R.S., O'Shaughnessy A.,
RA Rodriguez M.A., Shekher M., Kirchoff K.A., Baker J.P., Schutz K.,
RA Dedhia N.N., McCombie W.R.;
RT "Genomic Sequence For Oryza sativa, Niponbare Strain, Clone
RT OSJNBA0061K21 From Chromosome 10, Complete Sequence."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA McCombie W.R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.

```

RA McCombie W.R.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Palmer L.E., Yu M., Spiegel L.A., Nascimento L.U., de la Bastide M.,  
 RA Vil D.M., Preston R.R., Huang E.N., See L.H., Shah R.S.,  
 RA O'Shaughnessy A., Rodriguez M.A., Shekher M., Kirchoff K.A.,  
 RA Baker J.P., Schütz K., Dedhia N.N., McCombie W.R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC016780; AAM08782.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 753 AA; 85824 MW; 34DFBC05BE47820E CRC64;  
  
 Query Match 46.7%; Score 43; DB 10; Length 753;  
 Best Local Similarity 53.3%; Pred. No. 76;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
  
 QY 1 MPAHLIQDDVSFPAP 15  
 :||| ||| | | | |  
 Db 705 VPASLEVDVDTSLPRW 719

Search completed: May 14, 2003, 15:30:54  
 Job time : 37 secs